

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 19:22:54 ; Search time 177 Seconds
(without alignments)
4486.776 Million cell updates/sec

Title: US-09-522-753-5

Perfect score: 13215

Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEFPKPLCSQYETLSDSE 2517

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_25:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_phase:**

10: sp_plant:**

11: sp_rodent:**

12: sp_virus:**

13: sp_vertebrate:**

14: sp_unclassified:**

15: sp_rvirus:**

16: sp_bacteriap:**

17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4149	31.4	1998	13	Q8QFM0
2	4146	31.4	1988	13	Q8QFL9
3	4043	30.6	2343	4	Q7Z516
4	4009.5	30.3	2498	13	Q8Q378
5	3178	24.0	606	4	Q9BT61
6	3012	22.8	631	11	Q802V7
7	2184	16.5	914	4	Q86YX0
8	2155	16.3	527	13	Q7SVR4
9	2091.5	15.8	1724	11	Q8CHH6
10	1839	13.9	1399	11	Q8VDE8
11	1687	12.8	550	4	Q86W58
12	1272	9.6	1047	4	Q9NS20
13	1172.5	8.9	959	4	Q86YX1
14	1077	8.1	3604	5	Q9VYX0
15	1075	8.1	3469	5	Q9U412
16	827.5	6.3	302	13	Q7ZUE9

Q86YY2 homo sapien
Q8bk32 mus musculus
Q95ym2 procambarus
Q9w596 drosophila
Q76891 drosophila
Q81qa6 drosophila
Q86bh2 drosophila
Q9bz80 homo sapien
Q9nhx6 drosophila
Q8t9n4 drosophila
Q9vsk5 drosophila
Q95875 homo sapien
Q9p3j0 neurospora
Q9hcl7 homo sapien
Q96qc6 homo sapien
Q9ug35 homo sapien
Q7tsc1 mus musculus
Q8bt18 mus musculus
P70670 mus musculus
Q9vz30 drosophila
Q9z1r1 mus musculus
Q8wz42 homo sapien
Q83460 rattus norv
Q88737 mus musculus
Q10465 homo sapien
Q9n3y8 caenorhabdi
Q81q71 drosophila
Q88778 rattus norv
Q7xxn1 oryza sativ
Q9una8 homo sapien
Q9n541 caenorhabdi
Q61479 mus musculus
Q43161 homo sapien
Q9upa5 homo sapien
Q7yzhl drosophila
Q9wu02 drosophila
Q9y520 homo sapien
Q9c0a3 homo sapien
Q9vh10 drosophila
Q9p2r6 homo sapien
Q7yr40 pan troglod
Q9y519 homo sapien
Q9y3s1 homo sapien
Q9vc00 drosophila
Q8cf91 mus musculus
Q8cf92 mus musculus
Q8w966 homo sapien
Q81sf6 caenorhabdi
Q81sf7 caenorhabdi
Q9p6t1 neurospora
Q61461 rattus norv
Q57580 gallus gall
Q81q87 drosophila
Q15038 homo sapien
Q81sf5 caenorhabdi
Q9hec9 neurospora
Q9bx49 homo sapien
Q92954 homo sapien
Q9vid9 drosophila
Q75046 homo sapien
Q86z06 homo sapien
Q96qc2 homo sapien
Q9vwc0 drosophila
Q14676 homo sapien
Q17343 caenorhabdi
Q17490 caenorhabdi
Q60382 homo sapien
Q9w3z0 drosophila
Q15021 homo sapien
Q8ou37 mus musculus
Q95ku4 canis faml
Q966v1 drosophila
Q9v5j0 drosophila

90	405.5	3.1	1400	11	Q9ESU6	Q9esu6 mus musculus
91	405.5	3.1	1400	11	Q8VHF8	Q8vfh8 mus musculus
92	405.5	3.1	1480	10	Q9LIE8	Q9lie8 arabidopsis
93	405	3.1	2382	4	Q9H4A3	Q9h4a3 homo sapien
94	404	3.1	1274	5	Q20007	Q20007 caenorhabdi
95	403	3.0	1110	13	Q91255	Q91255 petromyzon
96	401	3.0	2656	5	Q9NE92	Q9ne92 leishmania
97	400.5	3.0	1118	5	Q9VR13	Q9vr13 pseudophila
98	400.5	3.0	3084	12	Q8UZ11	Q8uz11 pseudorabla
99	396.5	3.0	1464	4	Q8N473	Q8n473 homo sapien
100	396.5	3.0	2936	5	Q9NKP7	Q9nkp7 leishmania
RESULT 1						
Q8QFM0						
ID	Q8QFM0	PRELIMINARY;	PRT;	1998	AA.	
AC	Q8QFM0;					
DT	01-JUN-2002	(TREMELrel. 21, Created)				
DT	01-JUN-2002	(TREMELrel. 21, Last sequence update)				
DT	01-OCT-2003	(TREMELrel. 25, Last annotation update)				
DE	SC:B271M17.2.1	(Novel protein similar to mouse silencing mediator of				
DE	retinoic acid and thyroid hormone receptor (SMRT)) (Fragment).					
GN	SC:B271M17.2.					
OS	Brachydanio rerio (Zebrafish) (Danio rerio).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;					
OC	Cyprinidae; Danio.					
OX	NCBI_TaxID:7955;					
RN	[1]	SEQUENCE FROM N.A.				
RP	Laird G.;					
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.					
RC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).					
CC	-1- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.					
DR	EMBL; AL590153; CAD2407.1; -					
DR	GO; GO:0005634; C:nucleus; IEA.					
DR	GO; GO:0003677; F:DNA binding; IEA.					
DR	InterPro; IPR001005; Myb DNA binding.					
DR	Pfam; PF00249; myb DNA-binding; 2.					
DR	SMART; SM00717; SANT; 2.					
DR	PROSITE; PS0090; MYB_3; 1.					
KW	Nuclear protein.					
FT	NON TER					
FT	NON TER					
SEQUENCE	1998	AA; 216897	MW; 45860663	F3CF9EE	CRC64;	
Query Match						
Best Local Similarity 46.5%; Pred. No. 1.4e-191;						
Matches 1064; Conservative 222; Mismatches 560; Indels 440; Gaps 96;						
QY	384	NLEKQRLAVIPPMPLYADQQRIRKTNMGLMADPMKYKROVNMWSEOEKETEREK	443			
DB	1	NSEKQRLAVIPPMFLDAPQRIKTNMGLMDDPMKYKROVNMWSEOEKDTFERK	60			
QY	444	FMQPKNGLIASFLERKTVACVLYLTCKNENKSLVRSYRRGRKSQQQQQQQQQ	503			
DB	61	FIQHPKNFALIAFLERKTVACVLFYLTCKNENKTNVRSYRRGRSQNNQQAQQQ-	119			
QY	504	QQQQQQPMPSQQEERDEKEKEAEBEKEPEVNDKEDLLKKTDDTSGDNDKEB-	562			
DB	120	-----VNRNQEEKEDKEKEGEREDKAE-GDDKEDGK- -DDTSGDGEKEP	167			
QY	563	AVASKGRTANSQRRGRITRSMANSEFALTQOASALASMEINSSRWTEEMET	622			
DB	168	PVTYKGRNTANSRGRKGRVTRSGTNE- -QENTSPLGSELANLEWSSRWTEEMET	224			
QY	623	AKGLLEHGRNWSAIARMVSGKTVSQCKNFYFNYKKRQNDLEILQCHKLQKEKERNARK	682			
DB	225	AKGLLQYGRNWSAIARMVSGKTVSQCKNFYFNYKKRQNDLEILQCHKLQKEKERNARK	284			
QY	683	KKKAPAAASEAAPPVVEDEMEASGVSGNEEEMVEEAEALHASGNEVPRGCSGPATV	742			

DB	295	GK- -AQNEEASAPSAEEMEGSGVSGNEEAEQD- - - - -	GBGGA	324			
QY	743	NNSDTSGIPSPHT- - -EAKDTCQNGKPKPATLGAOPPGPPPTPRRTSRAPLETPA	799				
DB	325	NNSDTSLSPRSGDDSKVKEGSAAPK- - - - -	TAPNATS	- - - - -S	361		
QY	800	SEATCAPTPPPAPSPSAPPVVPVKEEKEETAAPVVEGEEQKPPAAELAVDTQKAE	859				
DB	362	SSST- - - - -	DKMAAA- - -ETGMDEKPKVKEGDGQAGVKQ	394			
QY	860	EPVSECTEAEAGPAKGDAAEATAAGALKAEKKEGGSGRATTAKSSGAPODSSSA	919				
DB	395	E-VKTEPGEGASEGDKPPALE-VREGKKEVKKKEHGG- - - - -	KSGA-HDSDSA	444			
QY	920	TCSADEVDEAGGDKVRLLSRPSSLITPTGPRANASPKPLDLKOLKQRAAALPPIQVT	979				
DB	445	TCSADEVEETESEKS- - - - -RPSLLSFSDHGVSSQAQKPLDLKOLKQRAAALPPI- - -	496				
QY	980	KVHEPPREDAPTKDPAPPQPQL- - - - -QPESDAPQOP- - - - -	1015				
DB	497	-LPEVSGRDNREGLPGKQAMPDALALYQQQITMAHESQEAQKQQQQQPTKSHQYPAAGD	555				
QY	1016	GSSPRKGSRSAPPADKE- - - - -AFA- -AEAKLPGDPPCWT	1051				
DB	556	SGSPRVTRSPG- - - - -ISDREERELEGNRLQALLHLGHSRAPSLADGKQKPGADDIRAL	614				
QY	1052	GLPPPV- - - - -PPREVIKASHAPDPSAPSPAPPGHPLPLGLHDTARFVLP	1106				
DB	615	GLGRVMSYLSFRDWAKFS- - - - -QFOLLHYNP- - - - -SLASTLQPDQDMAAVRPLHIPPP	667				
QY	1107	PLISSAKHPSVLERQIGAISQMSVOLHVPY- - -SEHAKAPVGPVTMGLPLMDPKKLAPF	1164				
DB	668	PLISSAK-PG- - - - -GSITQGTVPQLHNPSHSGHGVKVP-APGSLALSW-VDQRMWGF	718				
QY	1165	SGVKQQLSPRQAGPPESLIGVPTAQEASVLRGTALSGVPGSGITKIGISTPRVSDSAT	1224				
DB	719	PVVKQQLSPRGASSSSSSQAENLSTQDSSTSG- - -PMPAVQGGSIKIGICTRVHODSPIS	777				
QY	1225	YR- -GSITHGTPADVLYKGTITRIIGBDSPSRLDRGRDPSLPGKHVYIEGKHVLSYEGG	1283				
DB	778	YRGSGITQGTADVLYKGTITRIILNEDSPSR-ERSREDTPSKGHVYIEGSHILSFERR	836				
QY	1284	MSVTQSKEDGRSSSGPPHETAPKETYDMMEGRVGR- - -AISSASIEGLMGRAIPPERH	1340				
DB	837	PN- - - - -SKEEGCGDVG- - - - -AGMKSYDMMEVGISRRVPPIDPLSEGLIRAMPDPRD	887				
QY	1341	SPHILKEQHIRGSITQGISPRVYEAQEDYLREAKLLKREGTTPPPPPPPSRDLTEAKTQ	1400				
DB	888	SPH- - - - -POLIRGSIQGIIPR- - - - -DDCORREVQMKREGSPSPRGHS-TDALKSR	936				
QY	1401	ALGPLKLKPAHEGLVATVKEAGRSIHEI PREELRHTPELAPRLKEGSIQTGTPLYD	1460				
DB	937	- - - - -SHESMVTVKEGGRSIHLIP- - - - -PEGVVIQKP-KEGSIQTGTPLKQE	979				
QY	1461	TGASTTGKXHDVRLSGSPGRTFPPVHPLDVMADARALERACYEES- - -LKSRRPCTASSG	1519				
DB	980	PPGS- - - - -SKRHDVRSIIASSPR- - - - -PVSGLPSHLDPRGSDRAREYEEAGKSRPSAVVSAP	1033				
QY	1520	GSIARGAPVIV- - -BELGKPRQSPLTYEDH- - -GAPPAGHLPRGSPVTMREPTPLQSGSL	1574				
DB	1034	SSLSRVSPALSGQSSRPSPHSPVGYEDHKRSSYPPSPH- - -RASPLSGRNSQAHG- - -	1089				
QY	1575	SSSKASQDRKLTSTPREI- - -AKSPHSTVPEHHHPHPISPYHLLGVSGVDLYRSHIP- - -LA	1631				
DB	1090	SSKPOQQRKATPTPREMGSTKSP- - -LAVGEHTAN- - -LAFERIL-GL-GPDMYRSOMSHLT	1144				
QY	1632	FDPTISIPRIGIPLDAAAAYLPRHLAPNPTYPHLYPPYLIRG- - -YPTTALENROTIINDYI	1690				
DB	1145	FDAAALPRGIPID- - -SAYVLPRLHAPGPGYPHPYPPVYIRGCFPETALENROTFENDYI	1202				
QY	1691	TSQQMHHTATA- - -MAQRADMLRGLSPRESSIALNYAAGPRGIIDLSQVPHLPVLVPPTP	1748				

Db 156 LYNQPSDTKVHENIKTNQVMRKULILFFKRNHARKQBQIKORYQLMEAWBEKKVDR 215
Qy 316 TENNPRRAKESKVREYYEKOPFEIRKQRELQRMQSRVQSGSLMSAARSEHEVSEI 375
Db 216 TENNPRRAKESKREYYEKOPFEIRKQREQERFQ-RVQGAGLCSATIAHSBEHISEI 274
Qy 376 IDGLSEQNLKQOMROLAVIPMLYDADOORIKFIMNGLMADPMKYKYKDRQVMMWSEQ 435
Db 275 IDGLSEQNSBKQRLSVIPMMFDABQRRVKFIMNGLMBDPMKYKYKDRQVMMWTDH 334
Qy 436 EKETPREKPMOHKPNFGLIASFLERKTVACVLYLTKKNEYKSLVRRSY-RRRGKSO 494
Db 335 EKEIFKDFIOPKPNFGLIASFLERKSPDCVLYLTKKNEYKALVRMYKGRGRNQ 394
Qy 495 QQQQQQQQQQQQQPPRSQEEKBEKEAEKEEKEPEVENDKEDLLKEKTDTS 554
Db 395 QIARPSQEEKVEEKEE--DKAEKTEKKEEKKOBEKDEKESKENTKE--KOKIDGTA 449
Qy 555 GEDNDEKAVASKGRKTANSQRRKGRITRSMANSEBALTPOOS-----AEL 604
Db 450 -BETEERQATPRGRKTANSQRRKGRITRSMNEAAAAAASAAAAAATBEPPLPPPPPEP 508
Qy 605 ASMELNESSRWTEERMETAKGLLHGRNWSAIAARMVGSKTVSOCKNFYENYKKRONLDE 664
Db 509 ISTEPVTSRWTEERMEVAKGLVEHGRNWAIAKMWGTKEAQCKNFYNYKRNHJDN 568
Qy 665 ILQOHLKWEKERNARRKKKAPAAASEEAFPPVVEDEMEASGVSGNEBEMVEEABAL 724
Db 569 LLOQHKOQTSRKPREERDVSCQESVASTVSA-----QEDEDIAS-----NESENEDSEG- 619
Qy 725 HASGNEVPRGCSGPATVNNSDTESIPSPH-TEAAKTONGPKPATLGADEGPPGPP 783
Db 620 -----AENSDDTESAPSPPVEAVK-----PSED 643
Qy 784 TPRRTSAPTEPTPASEATGAPPPAP-PSPS-APPPVVPKEKEBETAAAPV---- 837
Db 644 SPENATSKGNTAPELE-----PTTETAPSTSPSLAVSTKPADESIVETQVNDISABT 699
Qy 838 -----BEGERQKPAABELAVDTGKABEPYKSECTEABEGPAGKDAEAAE 884
Db 700 AEQMDVDQOEHSAEBSVCDPPATK--ADSDVVEVRYPENHASKVEGDNTKERDLDR- 756
Qy 885 ATAEALKAKEKGGSGRAATKASSGA-----PQ-DSDSATCSADEVDEAGGDKNRLLS 939
Db 757 -----SEKVEPRDEDLVVAQINAOQRPQSDNDSSATCSADE--DVDGEPRQRMF 806
Qy 940 P---RPSLLTGTGPRANASQK--PLDLKOLKQRAAIPPI----- 976
Db 807 PMSKPSLLNPTGSILV--SSPLKENPLDLQLOHRAAVIPPMVSCTCPNIIPIGTFVSGYA 865
Qy 977 ----QVTKVHEPPREDAAPTKPAAPPQQNLOPESDAPQOGSSPRGKSRSPAPPADK 1032
Db 866 LYORHIKAMESALLE-----EQROEOIDLECRSSTSPCGTSKSP-----NR 909
Qy 1033 EAFAAEAQKLPDGPCTWTSGLFPFVPPPREVIKASHAPDPGSAFYAPGHPPLPLGLHDTA 1092
Db 910 E-----W-----EVLQAPAH-----QVITNLPEGVRLP-----TT 934
Qy 1093 RPVLPRPTIINPPLTSSAKHPSVLERQIGAISOG-----MSVLHVYPSEHAKAP-VG 1146
Db 935 RPTPPPPPLIPSSKTTTVASEK-PSFI--MGSGISOGTGTTLTSHNOASYTQETPKPSVG 991
Qy 1147 PVTMGLPLPMDPKKLAPPSSGVKQEQLSPRGQAGPPESLGVPTAEASVLRGTALGSVPGG 1206
Db 992 SISGLFRQOESAKSATLPYIKQEFSPRSQNSQPEGLLV-RAQHEGVVRGTA-GAIOEG 1049
Qy 1207 SITGIPSTRVPSDAITYRGSITHGTPA-----DVLVYKGTITRIIGEDSPSLDRG 1258
Db 1050 SITRGTPTSISVESIPSLRGSITQGTALPQTGIPTEALVKGSISRMPIEDSSP--EKG 1107
Qy 1259 REDSLPKGHVYIEGKGHVLSYEGMSVTCQSKEDGRSSSGPPHETAPKTYDWMGVRV 1318
Db 1108 REEAASKGHVIEGKSGHLSYDNIKNA-----REGTRSPRTAHEISL-KRSESVEGNI 1161

Qy 1319 GRAIS-----SASIEGLMGRAIPIPERHSPH-HLKEQHHRIGSITQGIPIRSVYEAQEDYL 1371
Db 1162 KQGSMSRESPVSAPLEGLICRALP--RGSPHSDLKERTVLSGSIQGTPTATTESFEDGL 1219
Qy 1372 BREAKLLKREGTPPPPPSRDLTETAYKTQALGPLKLKPAHEGLVATVKEAGRSIHEIPRE 1431
Db 1220 -KYPKQIKRES-----PPIRAFEGAI-----TKGKP-YDG-ITTIKEMGRSIIHEIPRQ 1264
Qy 1432 EL-----RHTPELPLAPRLKEGSIQGTPLKYDTGASTTGSKGDVRSILGSPQRTFPP 1486
Db 1265 DILTQESKTEVQVSTRPIIEGSIQGTPIKFDNN-SQSAIKKNVKSILITGPKSLERG 1323
Qy 1487 VHPLDVMAD-ARALERACYE-----ESLSRPGTASSGSGSIARGAPVIVPELKGPROSP 1540
Db 1324 MPPELIVENIKVVERGYEDVKAGETVRSRHTSVVSGSPVLST--LHEAPKAQLSP 1380
Qy 1541 LTYEDHGA-----PFAGHLPRGSPVTMREPTPRLOEGSLSSKA-SQRKLTSTPRE--- 1591
Db 1381 GIYDDTSARRTPVSYQNTMSRGSPPMNR-----TSDVSNKSTNHERKSTLTPTQRES 1433
Qy 1592 -IAKSPHSTVPEHHPHIPISPYEHLRLRGVSDLYRSHIPLAFDPTSIPRGIPLDAAAAYV 1650
Db 1434 IPAKSPVGVDPVUSH--SPFDPHHRGSTAGETVRSHTLHP-AMFPHRALDPAAYVL 1490
Qy 1651 LPRHLAPNTYPHYLPYPIRGPDTAALEN-RTIINDYITISQOMHNTATAMAAQADM 1709
Db 1491 FORQLSPTTPGYPSQVLY-----AMENTROTILNDYITISQOMQVNL-----RPDV 1535
Qy 1710 LRGLSPRESSIALNYAGPRGIIDLQVPHLPVLVPTPGTPTATMDRLAYLPTAPQPPS 1769
Db 1536 ARGLSPREQPLGLPYPA-TRGIIDLTMNP--TILVPHPGGTSTPPMDRITVPGTITFP 1593
Qy 1770 SR-HSSSPSPGPGTHLTKTPTTSSSERDRDRDRDREREKSILSTTTTVEHAPIWR 1828
Db 1594 PRPTNSMSGCHTHL---AAAASAEKEREKEREKERER---IAASDOL-----YLR 1643
Qy 1829 PGTEQSSSGSSGGGSSSRSPASHAHQHSPISPRTQDA-LQORPSVLHNTMGKII 1887
Db 1644 PGSEQ-----PGRPGSHGVYRSPSP-SVRTQETMLQORPSVFGTNGTSVI 1688
Qy 1888 TAVPSPKTVLURSTSTSPV-----RPAATFPATHCPGLGTLGTVPTLMBPVLPLK--- 1940
Db 1689 TPLDPTAQLRLMPLPAGGPSISQGLPASRYNTAADA-LAALVDAASAPOMDVSKTESK 1747
Qy 1941 -EAPRVARPERPRAD-----TGHAFLAKPPARSGLEPASS- 1974
Db 1748 HEARLEENLSRGAASVSEOOOLEKQTLVEKRSVQCLYTSAPPSPGK-----QPHSV 1802
Qy 1975 --PSKSEPRPLVPVSGHATIAITPAKN--LAPHASPDPPAPPASADPHREKTSQKP 2030
Db 1803 VYSEAGKDKGP--PPKSYREBELTRGKTTITAAANFIDVITRQIASDKDARERSQSSD 1860
Qy 2031 PSIOELELRSLGYGSSYSGEVPSVPSLTHDKGLPKHLELKDHLGELRKPQ 2090
Db 1861 SS-----SSLSSHRYETPSDAIEVISPASSPAPPOEKLOTYQPEVVKANQENDPTRQY 1914
Qy 2091 PGPVKLGEAAHLPHLPSPESQSSPPLLQTAG-----VKGHORVVTLAQHSIV 2142
Db 1915 EGP-----LHRYR---QQESPQQQLPPSSQAEGMQVPRTHRLITLADHIQOI 1962
Qy 2143 ITQDYTRHH-----PQQLSAPLAPLYSFPGA--SCPVLDIRRPPSDLYLPPPD-----H 2190
Db 1963 ITQDFARNOVSSQTPOO--PPTSTFQNSPALVSTPV--RTKTSNRYSPESQAQSVHH 2016
Qy 2191 GAP-ARSGPH-----SEGGKRSPEPNKTSVLGGEDGIEPVSPPEGMTPEGHSRAVPL 2244
Db 2017 QRPSRVSPELNDKRSRPGKSPERSHV---SSEPYEIPSPQ--VPVVEHKQDSLLL 2071
Qy 2245 LYRGEQTEPSRMGSKPGNTSQPPAPESKLTESNAMVSKKQEIKNKANTHNERNEFY 2304
Db 2072 LSQRBAPAEQRNDARSPGISYLPSPFTKL-ENTSPMVKSKQEIFRKLJNSGGGSDM 2130

Db 476 LGANAFNPLNASLPAAMPITTTADGRSDHALTSPGGGKAKVSGRPSRKAASPAAGL 535
Qy 2422 ASGDRPSSVSHSEGCNRTPLTNRWEDRPSSAGSTPPYNNPLIMRLQAGVWASPPP 2481
Db 536 ASGDRPSSVSHSEGCNRTPLTNRWEDRPSSAGSTPPYNNPLIMRLQAGVWASPPP 595
Qy 2482 PGLPAGSGLAGPHHAWDEEPKPLLCQYETLSDSE 2517
Db 596 PGLAAGSGLAGPHHAWDEEPKPLLCQYETLSDSE 631
RESULT 7
Q86YV0 PRELIMINARY; PRT; 914 AA.
AC Q86YV0, 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE NCOR isoform b.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Yu L.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF303586; AAC32942.1; -;
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF00249; myb DNA-binding; 2.
DR SMART; SM00717; SANT; 2.
DR PROSITE; PS50090; MYB 3; 1.
SQ SEQUENCE 914 AA; 103812 MW; 619057D2BD26480C CRC64;
Query Match 16.5%; Score 2184; DB 4; Length 914;
Best Local Similarity 51.4%; Pred. No. 3.1e-97;
Matches 490; Conservative 110; Mismatches 208; Indels 146; Gaps 26;
Qy 137 KDRSLTKLE-PVSPSPPHDPELPLPPLPSKEELIQMDRVDRDIARVEQOILKLLK 195
Db 36 QDPAFGGKEAPSPISGQCGDDQNASPSKLSKEELIQMDRVDRDIARVEQOILKLLK 95
Qy 196 KQOLEEAAKPEPEKPVSPPIESKHSRLVQIYDENRKAEEAAHRLLEGLPVELP 255
Db 96 KQOLEEAAKPEPEKPVSPPIESKHSRLVQIYDENRKAEEAAHRLLEGLPVELP 155
Qy 256 LYNQPSDTROYHENIKNOAMRKLILYFKRRNHARKQKQFCORYDQLMLEAKKVER 315
Db 156 LYNQPSDTKVHENIKTNQVNRKLLILFFKRRNHARKQKQFCORYDQLMLEAKKVER 215
Qy 316 IENPRRAKESKVREYKQFPEIRKQELQERMQSVQRCGLSMSAARSEHSEI 375
Db 216 IENPRRAKESKVREYKQFPEIRKQELQERMQSVQRCGLSMSAARSEHSEI 274
Qy 376 IDGLSEQENLEKONROLAVIPMLYDADQORIKFINMGLMADPMKYKQDVNMWSEQ 435
Db 275 IDGLSEQENLEKONROLAVIPMLYDADQORIKFINMGLMADPMKYKQDVNMWSEQ 334
Qy 436 EKETPREKFMQHPKNFGLIASFLERKTVAECVLYYLTNNENYKSLVRASY-RRRSQ 494
Db 335 EKETPREKFMQHPKNFGLIASFLERKTVAECVLYYLTNNENYKSLVRASY-RRRSQ 394
Qy 495 QQQ 554
Db 395 QIARPSQEKVEEKEE--DKAEKTEKKEEKKDEKEDKENTKE---KDKIDGTA 449
Qy 555 GEDNDKEAASVSKGRKGRITRSWANEANSEEAITPQOS-----AEL 604
Db 450 -EETEEQATPRGRKGTANSQRRKGRITRSWANEANSEEAITPQOS-----AEL 508

Qy 2340 MGLEAIRKALMGKYDQWEESPPLSANAFNPLNASLPAAMPITTTADGRSDHTLTSPG 2399
Db 429 MGLEAIRKALMGKYDQWEESPPLSANAFNPLNASLPAAMPITTTADGRSDHTLTSPG 488
Qy 2400 GGAKEVSGRPSRKAASPAAGLSDGDRPPSVSHSEGCNRTPLTNRWEDRPSSAGS 2459
Db 489 GGAKEVSGRPSRKAASPAAGLSDGDRPPSVSHSEGCNRTPLTNRWEDRPSSAGS 548
Qy 2460 TPRFYNPLIMRLQAGVWASPPPGLPAGSGLAGPHHAWDEEPKPLLCQYETLSDSE 2517
Db 549 TPRFYNPLIMRLQAGVWASPPPGLPAGSGLAGPHHAWDEEPKPLLCQYETLSDSE 606
RESULT 6
Q80ZV7 PRELIMINARY; PRT; 631 AA.
AC Q80ZV7, 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Similar to nuclear receptor co-repressor 2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC047524; AAH47524.1; -;
DR GO; GO:0004872; F:receptor activity; IEA.
KW Receptor.
FT NON_TER
SQ SEQUENCE 631 AA; 66869 MW; CC1F52630A984D6E CRC64;
Query Match 22.8%; Score 3012; DB 11; Length 631;
Best Local Similarity 90.6%; Pred. No. 2.2e-137;
Matches 576; Conservative 11; Mismatches 39; Indels 10; Gaps 6;
Qy 1887 ITAVEPSKPTVL---RSTSTSPVPAATFPFATHCPLGGLDGVYPTLMEPVLKPAP 1943
Db 1 VTSVEPTGTVLRWARSTSTSPVPAATFPFATHCPLGGLDGVYPTLMEPVLKPAP 60
Qy 1944 RVARPERADTGAFAPARSGLEPASPSKSEPEPLVPVSGHATARTPAKXLA 2003
Db 61 RVARPERADTGAFAPARSGLEPASPSKSEPEPLVPVSGHATARTPAKXLA 117
Qy 2004 PHASDPAPPASADPHREKTQSPFSGIQLERSLGYH-GSSYSEGVPEVPSVSP 2062
Db 118 PHASDPAPPASADPHREKTQSPFSGIQLERSLGYH-GAGYSPDGVPEVPSVSP 176
Qy 2063 SLTHDKGLPKHLEBKSHLEGELRPKQGPVKLGEEAAHLPRLPESQSSPLLOT 2122
Db 177 SLTHDKGLPKHLEBKSHLEGELRPKQGPVKLGEEAAHLPRLPESQSSPLLOT 236
Qy 2123 APGVKHQVVTVAQHLSEVITQDYTRHHPQOLSAPLPAPLYSPGASCFVLDRPPSD 2182
Db 237 APGVKHQVVTVAQHLSEVITQDYTRHHPQOLSAPLPAPLYSPGASCFVLDRPPSD 296
Qy 2183 LYLPPPDHGAAPARGSPHSEGGKSPFNKTSVLGGEDGIEPVSPGTEPHSAYV 2242
Db 297 LYLPPPDHGAAPARGSPHSEGGKSPFNKTSVLGGEDGIEPVSPGTEPHSAYV 356
Qy 2243 PLLYRDEQTEPSRMGSKSPGNTSOPAPFSKLTESAMVSKQOEINKLNTNRNEP 2302
Db 357 PLLYRDEQTEPSRMGSKSPGNTSOPAPFSKLTESAMVSKQOEINKLNTNRNEP 415
Qy 2303 EYNISQPGTEFNNPAITGTGLMYRQVQEHASTNMGLEAIRKALMGKYDQWEESP 2362
Db 416 EYNISQPGTEFNNPAITGTGLMYRQVQEHASTNMGLEAIRKALMGKYDQWEESP 475
Qy 2363 LSANAFNPLNASLPAAMPITTTADGRSDHTLTSPGGGKAKVSGRPSRKAASPAAGL 2421

QY	2439	CNRRTPNVRWEDRPPSAGSTPPYPNPLIMRLQAGVMASPPPLPAGSGPL--AGPH-	2495
Db	1646	YHRQTP--GWAWEDRPPSTGTFPPNPLTIRM-----LSSTPTQIACAPSAITQAAPHQ	1699
QY	2496	--HAWDEPKPLCSQYETLSDSE	2517
Db	1700	QNRWEREPAPLLSAQYETLSDSD	1723
RESULT 10			
Q8VDE8	PRELIMINARY; PRT; 1399 AA.		
AC	Q8VDE8;		
DT	01-MAR-2002 (TremBLrel. 20, Created)		
DT	01-MAR-2002 (TremBLrel. 20, Last sequence update)		
DT	01-OCT-2002 (TremBLrel. 22, Last annotation update)		
DE	RIP-13.		
GN	NCOR1 OR RIP13.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Liver;		
RX	MEDLINE=97120602;		
RA	Seol W., Mahon M.J., Lee Y.K., Moore D.D.;		
RT	"Two receptor interacting domains in the nuclear hormone receptor		
RT	corepressor RIP13/N-CoR.;"		
RL	Mol. Endocrinol. 10:1646-1655(1996).		
DR	EMBL; L78294; AAL40135.1; -.		
DR	MGI; 1349717; Ncor1.		
SQ	SEQUENCE 1399 AA; 151393 MW; D42406CA592BEE20 CRC64;		
Query Match 13.9%; Score 1839; DB 11; Length 1399;			
Best Local Similarity 35.5%; Pred. No. 2.3e-80;			
Matches 557; Conservative 210; Mismatches 517; Indels 284; Gaps 72;			
QY	1057	VPPREVKASHPADPSAFSAPPCHPLPLGLHDTARVLPRLPPTISNPPPLISSAKHPS	1116
Db	8	VPP--VLQAPH---QVITNLPEGVLP-----TTRTRPPPPPLIPSKTTVASEK-PS	55
QY	1117	VLERQIGAISQ-----MSVLHVVPYSEHAKAPGVFTVMGLPMDPKKLAPFGSVKQEQ	1171
Db	56	FI--MGGSISQCTPTGYLSSHNOAYPOEAPKPSVGSISLGLPRQOESTKAAPLTIKQEE	113
QY	1172	LSPRQAGPPSELGVPTQAQASVLGRTALGSPVGGSIKGPSTPSTVPSDSAITYRGSITH	1231
Db	114	FSPRSQSQPEGLLV-RAQHEGVWRGTA-GAVQEGSITRGTPASKISVETISSLRGSITQ	171
QY	1232	GTPA-----DVLKGTITRI-IGEDSPSLDRGREDSLPKGHVYEGKKGHVLSEY	1282
Db	172	GTPALPQAGITPEALVGPVSRMPLEESSPEKV---REEAASKGHVYEGSGHLLSY--	226
QY	1283	GMSVTQCSKEDGRSSGGPPHETAAPKRTYDMMEGRVGRGAISSASIEGLMGRAIPPERHSP	1342
Db	227	-----DSLICRALP--RGSP	239
QY	1343	H-HLKEQHIGSITQGI PRSVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTOA	1401
Db	240	HSDLKERTVLSGSINQGTFRATAESFEDGL-KYPKQIKRES-----PPIRAFEGAI----	289
QY	1402	LGPLKLKPAHGLVATVKEAGRSIHEIPREL-----RHTPELPLAPPLKEGSTTQCTP	1456
Db	290	---TKGKP-YDG-ITTIKEMGRS IHEIPRODILTOESRKTPEVVQSTRTPIEGSISQGTTP	344
QY	1457	LKYDTGASTGSKKHVRSLSIGSPGRTFPFVHPLDVMAAD-ARALERACYE-----ESLKS	1510
Db	345	IKFDNN-SGQSAIKHNKSLITGFSKL--PRGMLDIVPENIKVVERGYEDVKAGEPVRA	401
QY	1511	PGTASSGGSIAGAPVIVPELKGPRQSPLTYEDHGA-----PFAGHLPRGSPVTMREP	1565

Db	402	RHTSVVSGSPVLST---LHEAPKAQLSPGLYDDSSARRTPVSVYQNTISRSGSPMMNR--	456
QY	1566	TPRLQEGSLSSK-ASQDRKLTSTPRE---IAKSPHSTVPEHHPHPIPSPYEHLRLGVSG	1620
Db	457	-----TSDVSSKSASHERKSTLFTQRESIPAKSPVPCVDPVVSH--SPFPDHRSSAA	509
QY	1621	VDLYRSHPLAPDPTSIPRGIPLD-AAAAYYLPRHLAPNTPYPHLYPPYLIIRGYPDTAAL	1679
Db	510	GEVYRSHLPTLDP-AMPFHRALDPAAAAAYLLQRLSPTPGPYSQYLY-----AM	559
QY	1680	EN-RCTIINDYITSOOMHNTATAMAQRADMLRGLSPRESSIALNVAACPRGIIDLQVPP	1738
Db	560	ENTQTILINDYITSOQMQVNL-----RPDVTIRGLSPREQLGLPYPA-TRGIIDLTNMP	612
QY	1739	HLPLVLPPTPGTATAMDRLAYLPTAPQPFSSR-HSSSPLSPGGTHLTIKPTTTSSSRE	1797
Db	613	P-TILVPHAGGTSTPPMDRIYIFGTQVTFPPRPYNAASLSFGCHTHL---AAAASAERE	668
QY	1798	RDRDRDRDREREKSILTSTTTTVEHAP---IWRFGTEQSGSGSSGGGGSSSRPASH	1854
Db	669	REREREKERERERERERIAAAPADLYLRPGSEQ-----PCRPQSH	714
QY	1855	SHAHQHSIPSPRTQDA-LQORPSVLHNTGMKGIIITAVEPSKPTVLRSTSSPV-----RP	1909
Db	715	GYVRSPSP-SVRTQETILQORPSVFGTNGTSVITPLDPTAQLRIMPSPSGGSPISQGLP	773
QY	1910	AATFPPTHCPGLGGLDGVYPTLMBPVLLPKEAPR-----VARPERPRADTG	1956
Db	774	ASRYNTAADA-LAALVDAASAPQMDVSKTKESKEAARLEENLRSAVAVEQQOLEQ	832
QY	1957	HAFIAKPPARSGLPPASSPSKSGEPRLV-----PPVSGHATIAARTPAKN--LA	2003
Db	833	NLEVEKRSVQCVTSSALPSGKAQPHASVAVYSEAGKDGPCKRKYEEELRTRGTTITA	892
QY	2004	PHASPPAPPASADPHREKTQSKPFSIQEELRSLGYHCSSVSPGCVFVSPVSSPS	2063
Db	893	ANFIDVIITROIADKADARERGSSOSSS-----SSLSSHRYETASDAIEVISPASSA	946
QY	2064	LTHDKGLPKHLELDKSHLEGELRPKQPGVKLGGEAAHLPHLR-----PLPESOPSSP	2118
Db	947	PPQEKQAYQDMVKANOAEENSTQYEGP-----LHHYRQQQSPSPSQOQPLPP	997
QY	2119	LLQT--APGVKGHORVVTLAQHI SEVITQDYTRHHPOQLSAPLAPLYSFPQA--SCPVL	2174
Db	998	SSQSEGMQVPRTHRLITLADHI CQIITQDFARN--QVPSQPSTSTFTQSPSALSSTPV-	1054
QY	2175	DLRPPSDLYLPDPD-----HGAPA-RGSPHS-----EGKRSPEPNKTSVLGGGEGDTEP	2224
Db	1055	--RTKTSRYSPEQSQTVLHPRPGRVSPENLVDKSRGPKSPERSHI---PSEPYEP	1109
QY	2225	VSPEGMTEPG-HSRSAVYPLLXRDGEQTEPSRMGSKSPGNTSQPPAFPSKLTESNAMV	2283
Db	1110	ISPPQG---PAVHEKQDSMLLSQRGVDPQORSRSPGSIYLPSPFTKL-ESTSPMV	1165
QY	2284	KSKQEIENKLTNHNREPEYNI SQPTEIFNMPAITGTGLMTYRSQVQEHASTNMGLE	2343
Db	1166	KSKQEIFRKLNSGGGSDMAAQPCTEINLPAVTTSGAVSSRSHSFADPAS-NLGLE	1224
QY	2344	AIIRKALMGKYDQWEE-----SPLSANAFNPLNASLASLPAAMPDITAADGRSDHTLSP	2397
Db	1225	DIIRKALMGSPDDKVEDHGVVMHPV---GIMPGSASTSV-----VTSSEARRDGEPS	1276
QY	2398	GGG-GKAKVSGRPSRRKAKSPAPGLA--SGDRPPSVSVHSEGCNRRTPLTNRVWEDRP	2454
Db	1277	HAGVCKPKLINKSNRKSKSPI PQOSYLGTRPSPSVSVHSEGDTHROTP--GWAWEDRP	1334
QY	2455	SSAGSTPPPNPLIMRLQAGVMASPPPPGLPAGSGPL--AGPH---HAWDEEPKPLCSQ	2509
Db	1335	SSTGSTQFPNPLTIRM---LSSTPTQIACAPSAITQAAPHQNRWIEREPAPLPSAQ	1390
QY	2510	YETLSDSE	2517
Db	1391	YETLSDSD	1398


```
Qy 2073 HLELDKSHLGEIRPKQPGVKVLGGEAAHPLHLRPLPESQPSPLLOTAPG----- 2125
Db      : : : : :
Qy 600 QPEVVKANQAEENDPTROVEGP-----LHHYRP-----QOESPQOQLPSSQOAGMG 647
Db      : : : : :
Qy 2126 -VKGHORVVTLAHISEVITODYTRHH-----POOLSAPLAPLYSPGCA--SCPVLDR 2177
Db      : : : : :
Qy 648 QVPRTHRLITLADHICQIITODFARNQVSSOTPOQ---PPTSTFQNSPALVSTPV---R 701
Db      : : : : :
Qy 2178 RPPSDLVLPDPD-----HGAP--ARGSPH-----SEGGRSPPEPKNTSVLGGEDGIEPVS 2226
Db      : : : : :
Qy 702 TKTSNRYSPESQAQSVHHORPGSRVSPENLVKSRGSRPGKSPERSHV---SSPEYEPIS 758
Db      : : : : :
Qy 2227 PPEGMTPECHRSR SAVVPLLYRDGTEQTEPSRMGSKSPGNTSQPPAPFSLKLTESNAMYKSK 2286
Db      : : : : :
Qy 759 PPQ--VPVVEHEKQDLSLLSQGAEPAEQRNDARSQSGISYLPSPFTKL-ENTSPMVKSK 815
Db      : : : : :
Qy 2287 KOEINKKLTNHRNEPEYNIQPGTEIFENMPLTGLMTYRSQAQVHASTNGLIAII 2346
Db      : : : : :
Qy 816 KOEIFRKLNSGGSDMAAAQPGTEIFNLPAVTTSGSVSRGSHFADPAS--NLGLEDI 874
Db      : : : : :
Qy 2347 RKALMGKYDQWEE-----SPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGG 2400
Db      : : : : :
Qy 875 RKALMGSDDKVEDHGVVMSQPMGV---VPGTANTSV-----VTSGETRREGDPSPHSG 926
Db      : : : : :
Qy 2401 G--KAKVSGRPSRRAKSPAP--GLASGDRPPSVSVHSEGDGNCNRRTPLTNRVWEDRPSS 2456
Db      : : : : :
Qy 927 GVCKPKLISKNSRKSPIPGQVGLGTERPSSVSVHSEGDYHRQTP--GWAWEDRPSS 984
Db      : : : : :
Qy 2457 AGSTFPFYNPLMRLOQVWASPPPGPLPAGSG--PLAGPH---HAWDEEPKPLCSQVET 2512
Db      : : : : :
Qy 985 TGSQFPYNPLTMRM---LSSTPTPIACAPSAVNQAAPHQONRIWEREPAPLLSAQVET 1041
Db      : : : : :
Qy 2513 LSDSE 2517
Db      : : : : :
Qy 1042 LSDSD 1046
Db      : : : : :

RESULT 13
Q86YV1 PRELIMINARY; PRT; 959 AA.
AC Q86YV1.
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE NCOR isoform c.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Yu L.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF303585; AAC32941.1; -.
SQ SEQUENCE 959 AA; 104500 MW; 35F97ED97CFB1CDD CRC64;

Query Match 8.9%; Score 1172.5; DB 4; Length 959;
Best Local Similarity 35.6%; Pred. No. 2e-48;
Matches 366; Conservative 127; Mismatches 355; Indels 181; Gaps 48;

Qy 1573 SLSSSKA-SODRKLSTPRE----IAKSPHSTVPEHHHPHPISPYEHLLRGLVSGVDLYRSH 1627
Db      : : : : :
Qy 1628 IPLAFDPTSPRGIPDLAAAAYLPRHLAPNTTPHLYPPVILIRGYDPTALEN-RTI 1686
Db      : : : : :
Qy 85 LPTHLDP-AMPFHRALDPAAYLFQRLSPPTPGPSQVQLY-----AMENTRQIL 134
Db      : : : : :
Qy 1687 NDYITSOOMHNTATAMAQRADMLRGLSPRESSIALNVAAGPRGIIDLQVPHLPVLVPP 1746
Db      : : : : :
Qy 135 NDYITSOOMQVNL-----RPDVARGLSPREQLGLPYPA--TRGIIDLTMNPP--TILVPH 186
Db      : : : : :
```

```
Qy 1747 TPGPTATAMDRLAYLPTAQPFSSR--HSSSPLSPGGPHLTKPTTTSSSERERDRDRD 1805
Db      : : : : :
Qy 187 PGGSTPMDRITVPGTQITFPPRPYNASMSGPHLTHL---AAAAGAEREREREKE 243
Db      : : : : :
Qy 1806 RDRERESKILSTTTVEHAPIWRBCTEQSSGSSGSGGSSRPPASHAHQHSPISP 1865
Db      : : : : :
Qy 244 RERER---IAAASDL-----YLRPGSEQ-----PGRPGSHGVYRSPSP--SV 281
Db      : : : : :
Qy 1866 RTQDA--LQORPSVLHNTGMKGIIITAVBPSKPTVLRTSTSSPV---RPAATFFPPATHCP 1920
Db      : : : : :
Qy 282 RTQETMLQORPSVFGINGTSVITPLOTAQLRMPLPAGGPSISQGLPASRYNTAADA-- 340
Db      : : : : :
Qy 1921 LGGTLDGYVPLMBPVLLPK-----EAPVARPERPRAD----- 1954
Db      : : : : :
Qy 341 LAALVDAASAPQMDVSKTKESKHEAARLEENLRSAVSEQQOQLEQKLTVEKRSVQC 400
Db      : : : : :
Qy 1955 --TGHAFILAKPPARSGLPASP---PSKGSFPRPLVPVSGHATARTPAKN--LAPPHA 2007
Db      : : : : :
Qy 401 LYTSAPFSGKP-----OPHSSVVYSEAGDKGP--PPKSYEBELRTRGKTITTAANFI 453
Db      : : : : :
Qy 2008 SPDPAPPASADPHREKTKOSKPFISQIELELRSLGYHGSSYSPGVEPVSPVSSPSLTHD 2067
Db      : : : : :
Qy 454 DVIITRIASDKDAREGSSQSDSS-----SSLSSHRYETPSDAIEVISPASSPAPQOE 507
Db      : : : : :
Qy 2068 KGLPKHLELDKSHLEGELRPQPGVKVLGGEAAHPLHLRPLPESQPSPLLOTAPG-- 2125
Db      : : : : :
Qy 508 KLQTYQPEVVKANQAEENDPTROVEGP-----LHHYRP---QOESPQOQLPSSSQ 555
Db      : : : : :
Qy 2126 -----VKGHORVVTLAHISEVITODYTRHH-----POOLSAPLAPLYSPGCA--SCP 2172
Db      : : : : :
Qy 556 ABGMGQVPRTHRLITLADHICQIITODFARNQVSSOTPOQ---PPTSTFQNSPALVSTP 612
Db      : : : : :
Qy 2173 VLDLRRPPSDLYLPPDP-----HGAP--ARGSPH-----SEGGRSPPEPKNTSVLGGSDG 2221
Db      : : : : :
Qy 613 V---RTKTSNRYSPESQAQSVHHORPGSRVSPENLVKSRGSRPGKSPERSHV---SSEP 666
Db      : : : : :
Qy 2222 IEPVSPPEGMTEPGHRSR SAVVPLLYRDGTEQTEPSRMGSKSPGNTSQPPAPFSLKLTESNSA 2281
Db      : : : : :
Qy 667 YEPISPPQ--VPVVEHEKQDLSLLSQGAEPAEQRNDARSQSGISYLPSPFTKL-ENTSP 723
Db      : : : : :
Qy 2282 MVKSKKQIEINKLTNHRNEPEYNIQPGTEIFENMPLTGLMTYRSQAQVHASTNMG 2341
Db      : : : : :
Qy 724 MVKSKKQIEIFRKLNSGGSDMAAAQPGTEIFNLPAVTTSGSVSRGSHFADPAS--NLG 782
Db      : : : : :
Qy 2342 LEAIIKALMGKYDQWEE-----SPPLSANAFNPLNASASLPAAMPITAADGRSDHTLT 2395
Db      : : : : :
Qy 783 LEDIIRKALMGSDDKVEDHGVVMSQPMGV---VPGTANTSV-----VTSGETRREGDP 834
Db      : : : : :
Qy 2396 SPGGG--QKAKVSGRPSRRAKSPAP--GLASGDRPPSVSVHSEGDGNCNRRTPLTNRVWED 2452
Db      : : : : :
Qy 835 SPHSGVCKPKLISKNSRKSPIPGQVGLGTERPSSVSVHSEGDYHRQTP--GWAWED 892
Db      : : : : :
Qy 2453 RPSAGSTPPFPYNPLMRLOQVWASPPPGPLPAGSG--PLAGPH---HAWDEEPKPLCS 2508
Db      : : : : :
Qy 893 RPSSTGSTQFPYNPLTMRM---LSSTPTPIACAPSAVNQAAPHQONRIWEREPAPLLSA 949
Db      : : : : :
Qy 2509 QYETLSDSE 2517
Db      : : : : :
Qy 950 QYETLSDSD 958
Db      : : : : :

RESULT 14
Q9VYK0 PRELIMINARY; PRT; 3604 AA.
ID Q9VYK0
AC Q9VYK0.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG4013 protein.
GN SMR OR CG4013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

Mon Apr 19 08:15:08 2004

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclev J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M.D., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RN Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jaiswal M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Pfeiffer B.,
RA Paclev J., Paragov S., Park S., Patel S., Pfeiffer B.,
RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003490; AAF48195.2; -;
DR FlyBase; Fgn0024308; Smr.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002086; Aldehyde dehydr.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF00249; myb DNA-binding; 1.
DR PROSITE; PS00687; ALDEHYDE DEHYDR. GLU; 1.
SQ SEQUENCE 3604 AA; 378155 MW; E7563A180CID546B CRC64;
Query Match 8.1%; Score 1077; DB 5; Length 3604;
Best Local Similarity 21.0%; Pred. NO. 4.4e-43;
Matches 713; Conservative 356; Mismatches 1083; Indels 1238; Gaps 143;
20 PPHSLSYFVQIARTHTDVGLEYQH--HSRDYASHLSPGSIIPQORRRPS----- 67
561 PPH-----PQQ-----QHQQQQQQQQQHOHQHQH--QRSSPAQVQQQQQQQWNSRQSHNDMC 612
68 -----LLSEFQPGNERSQELHLRPEHSYSLPELGSEMFIESKRPRLLELDDP 116
613 QVVVTPMGQMLKVELTLPQQQKQHQHQHQHQ-----QQQGRSQ-----PWSMSMTV 662
117 LLRP-----SPLATQPGASEDLTKDRSLTGKLEPVSPPSPHTDP 158
663 VSPGVCTVTTTAGLSASHSGSGNGVAAGLTGNTGTSASTAYHPQVEAISITLP--SS 720
159 ELELVPRLSKELIQNMORVDREITWVPOQISKLKKQOQLEEA--AKPPEKPVSV- 215
721 SIERGRTSAKEDLLMQIOVDNEIKSAFTTWTLEKESLMEEAALAKEQRAAKELND 780
216 -----PPPIESKHS--LVQIYIDENRKAEEAHLLEGLPQVE-----LPLYN 258
781 NNNDQPMVELSWKSWLAEKIYAANKTKAQASHMLQNAADDESSPGSVAGRPWPLYN 840
259 QPSDTQYHENIKINQA--MRKKLILYFKRNHARKQWKQFCORYDQLEALEKKVERIE 317
841 QPLDVEALAMLIHQHSQIRAPLLHLIRKLUKAERWAHNOQLVEKYTKDQADWQRCERME 900
318 NNPRRAKESKVRVEYKQPEIRKQRELMQSRVQSGSLMSAARSEHEVSEIID 377
901 ASAKKAKAREAKNREFFKVFTELKQREDKERF--NRVGR-----IKSEADLEIMD 951
378 GLSEQENLEKOMQLAVIPMLYDADQORIKFINNGLMADPMKVYKDRQVMMWSEQEK 437
952 GLQSALEDKMRSYAVIPPLMHDARQRCAYHNENGLIEDMVAVHQORALANNWTAGEK 1011
438 ETFREKEMQPKNFGILTASFLERKVAECVLYVYTKNENYKSLVRSYRRGKSOQQQ 497
1012 ETFKEKYLQHPKNGFALASLDRKSPQDCVRYVYLSKKTENYKQLLRKSRQRTSSRNPA 1071
498 QQQQQQ-----QQQQQQQPMPRSS-----QBEKKEKEKEKE----- 528
1072 KAQAQPCQIIDSMTTGTMTLQREQQQKSGRSSAVERAERAEAAERVAEKAADA 1131
529 ---AKSEERKE----- 537
1132 AKAAESAERAKAATAKAVEATAAGEKVAKAAAAAATAATTATTSSSSSSSSA 1191
538 -----VENDKEDLLK-----EKTDDTSG----- 555
1192 SSASTASSTASPATLACIADKTDAGKTASAKNAATAGGPTATGTPTAATTATATA 1251
556 -----EDNDEKEAVAKGRKGTANSQGRKGRITRSMANEANSEEAITPQQA-- 602
[4]

QY 773 ---LGADGPPPG-----PPTPRRTSRA-----PIEP 796
Db 1468 ISLPLLDGSGNMVANEIILALDKDKLASCFCVKAECACRTRPLKKGROQQYCIDPET 1527
QY 797 TPAS-----EATGAPTPPPAPSPSAPPV-----VPKEEBETAAAPVBE-- 839
Db 1528 IPAGARVCNSQCKSVRSRYNCPCLPTCPNPKDRAQLRNIPSLRFE-----LAEVRDPL 1583
QY 840 -GEBQKPPAA-----EELAVDTGKA-----EBPVKSECTEABE 872
Db 1584 MAEFQIPPHATRCACSACILMRIRRLKDPOLNLTGSSGGAGSGGDETDVSTSSCDERE 1643
QY 873 GPAKGDAABABATAGLAKAE-----895
Db 1644 G---GSDTASVESPEN---LORHKSILTMVKQOQOQOQOQOQOQOQOQOQOQOQO 1698
QY 896 -----KEGSGRATTA-----KSSGAP 912
Db 1699 PAPQOQKSGSRGDDQGTPLIITPTRMSSKSGSGAQTAGDNERLLPPAAGQAPKKQKTS 1758
QY 913 QDSDSATCSADEVDE-----AEGDKNRLSPRP---SLLTPTGDP 951
Db 1759 EBYDSSATETADEENENSPANQSPKVLPHGCHGHGHANNVAGLQPPVANMGTGGVQ 1818
QY 952 RANASPOK---PLDLKQLKQRAAIPPIQ---VTKVHEPPREDAAPTKPAAPPQNQLQ 1006
Db 1819 PGGAAGQOVNGPISM-----RREAIVNVQDCVFSVIERSLKHGP-----Q 1859
QY 1007 PESDAPQOPG-----SSPRKSRSPAPPADEAFEAQAOKLPGDPPCWTSGLPPV 1057
Db 1860 PKGQO 1904
QY 1058 PPREVIKASHPADPSAFSY-----APP---GHPLPLGLHDHTARPLPRPTISNP 1105
Db 1905 ERKELTVREYRQDPGLKQO 1962
QY 1106 PPLISSAKHPSVLERQIGAISQMSVOLHVPYSEHAKAPGVPTVGLPLMDPKKLAPFS 1165
Db 1963 PP-----AHP-LTPTSIGCAGNNGTSDSLATLSVNMHGMVGIHGPMAHASSAGGI 2016
QY 1166 GVKQEQOLSP---RGQAGPESIGVTAQEA5VL-RGTALGVPGGSIKIGISTRVPSDSA 1222
Db 2017 GVDKATITPVVKSSGSGSKSGGSGASHSTATPETIYNVPVAHPQORGIPP---PSOHS 2073
QY 1223 I--TYRGSITHTPADVLYKGTITRI--IGEDPSRLDRGRD5LPGKHVIEYEGKKGHVLS 1279
Db 2074 VHPAHPSTQHPAHPQSHSHQHTQLQVPBEPQTLDSLICKPRDGHSPHTG-----2126
QY 1280 YEGGMSVTQCSKEDGRSS-----GPPHETAAPK---RTYDMMEGRVGRAISSA-----1325
Db 2127 --AGSSSSGSGSGSGSSDRHHGPPPPPTMSKHIIVRSGMVYRGDTVTPVSLAAPSSVLY 2184
QY 1326 ---SIEGLMGRAIIPPERHSPHLKEQHIRSGITQGIIPRSVTEAQ-----EDYLREAK 1376
Db 2185 PTRSVKSIGGGVVP-----GVLPGVPGIALYLQVPVPVPIISGGQO 2228
QY 1377 LLKEGEGTPP---PPPSRDLTEAYKTOALGPLKLAP--AHEGLVATVKEAGRSIHEIPREB 1432
Db 2229 LPPKAGQPPPAQPSRGVAKV-----PPKLSQQAHLHPSHGHSFSSQOQOQOQOQOQO 2281
QY 1433 LRHTPELAPRPL--KEGSIQTGTPLKYDTGASTTGSKHGDVRSLLIGSPGRTFPPVHPL 1490
Db 2282 QOQOQOQAAAAQQLLVKSGSIHGTG---ANSAQOQIIIVHAPATAAAPSLSFKPD- 2336
QY 1491 DVMADARALEACVYESLSK--RPGTASSG--GSIARGAPVIVP-----1531
Db 2337 ---GLVRQTTPEGVSGVPGGAGSGKHGSIQTGTPLMPPHLESKRPIYESYKS 2389
QY 1532 ---ELGKPRO--SPLTYEDHG-----APFAG-----1552
Db 2390 SQRHSPAQQCGNQOQLPPPPQOQSPQAPPPGOGYGVGVSSPYARSPFAGVVEQPOVLSTRQ 2449
QY 1553 -----HLPRGSPVT-----MREPTPLQEGSL- 1575

Db 2450 IVMHDIYTSOQMOQO 2509
QY 1576 -----SSKASQDKLTSTPREIAKSPHSTVP-----BHH-----P 1605
Db 2510 PAYGDKESAPRGPEVSSRASPADHVNSTP-----SPHRTPPQORQVIOHNTGSKPP 2564
QY 1606 HP-----ISPYEHLRGVGVLYRSHIPLAFDPTSIP-----RGIPLDAAAAYY 1650
Db 2565 SPAAPPPSRMHPPIYQY---APSGHDALASPDVAVQOQPLVPFSQKDKSPGSTAPQO 2621
QY 1651 LPRH---LAPNPTYPHLY-----PP---YLIRGYPTDALENQRTIINDYTSOQMHN 1698
Db 2622 VPGSGPLGSPPLPHAVGWVQAPPPPTAHDQRYDLTLHHHTLTVQOOI-AQQOQHYR 2680
QY 1699 TATAMARADMLRGLSPRESSLALNYAAGPRGIIIDLSQVPHLPVLVPTGPTATAMDRL 1758
Db 2681 SLNVAQO--VDMORQMDQAKVRMRHQHQVQOQOQOQOQOQOQOQOQOQOQOQOQO 2728
QY 1759 AYLTPAQPFSSRHSSPLSPGPGTHLTKTPTTTSSSERERDRDRDREREKSLTST 1818
Db 2729 ---EMQERMREDR-----REREREQERERERERERERERERERERERERER 2772
QY 1819 TTV-----EHAPIWRPGTEOSSGS--SGSSGGGSGSSSRPASHSHAHQHSISPRTQDAL 1871
Db 2773 RRVVAEREHDS--RRMERMFAGNVVTGSGGAGGSPS--PGQFLRA-----2815
QY 1872 QORF5VLHNTGMKGIITAVEPSKPTVLSTSTSPVRPAATFPFATHCPLGTLGDVYPT 1931
Db 2816 -----SVPETGPPRSIPDRERESYVYQAHGCGPADETP--GQLSA--QS 2855
QY 1932 LMPEVL---LPKEAPRVARPER--PRADTGHAFIAKPPARS---GLEPASSPSKGSBP- 1981
Db 2856 LIDAIKHEINRSNDATAGPGRFPFRPSFVHAPL--PPRGSGGGGTCTRSSPANVLHPM 2913
QY 1982 --RPLVPVSGHATITPAKNLAPHASPDPPA-----PPASADPHREKTQS 2028
Db 2914 YLRDLRQPLDGGAGSMLTAENNGKP--SSGSPSVINIDLQERISAAAAVAQOQOQOQ 2971
QY 2029 KPFSIQEILRSGLYGHSSYSPGEPVPSVSSPSLTHDKGLPKHLELDKSHLEGLRP 2088
Db 2972 APPSQSSQSRV--HQQLRTPTS--QSGGSAAPSQQIHTKSIT--FGELTOSITSDYGT 3026
QY 2089 KQPGFVXLGGEAAHLPLRP-----LPESOPSSSP-----LLQTAQVKGH-----2131
Db 3027 N-----PHLRPYMAYLOETQISILPDRWKQNRMQKAEANDHSQOQOQ 3072
QY 2132 VVTLAQHISEVITQDYTRHHQQLSAPLPAP-LYSFPG-----AS 2170
Db 3073 QHQOQOQHAQO 3132
QY 2171 CPVLD---LRRP-----PSDLYLPPP-----2188
Db 3133 TPGEDGNIIRMPQAVSPRKFNMHMLHVMTGTGAGEAQOFLPSRVVLPEORGTSPG 3192
QY 2189 DHGAPARGSPHSEGGKSPBN-KTSV-----LGGEDGIEPVSPPEGMTEP--GHSR 2238
Db 3193 GGGAPGAGGCGSGGATTIEKYVKTIAEVNRDDIGYKNTVEVTEDEVTDWVAHSH 3252
QY 2239 SAVY-----PLLYRDEQTEPS--RMGSKSPGNTSQPPAPFSKLTESNSA 2281
Db 3253 AAVHAAHVAHAAHVAHAAAMELQHRSEKPEPPPEISVSRKTPNQEVVDASGRRSAGSGV 3312
QY 2282 MVKSKQOEIKKLNTHNRPENISOPGTEIF-----NMPAITGTGLMTRYSAQVQSHA 2336
Db 3313 SV-----SVGANSHHSPYHPAAAAYAPSTYAFYFYSALNVPGAAG-----GLPHQ 3358
QY 2337 STNMGLEAIRKALMGKYDQWEESFPLSANAFNPNASASLPAAMPITAAOGRSDHTLTS 2396
Db 3359 PLQLAHQAV-----APP---GAF-----AKAKAAHALSE 3384
QY 2397 PG--GGGKAKVSGRPSRKAAPAGLASGDRPPSVSVHSEGDGNCNRRTPLTNRWEDRP 2454

Mon Apr 19 08:15:08 2004

Db 3385 LGAVGGVSLVGGSG-----GIAGG--PGGVS----- 3411
 Qy 2455 SSAGSTFPYPLIMLQAGVMAAPPPGCPAGSGPLA-----GPHH----- 2496
 Db 3412 -----VG-----GVPGGGPGSGGGGGGGHNSSSQASAAVAAA 3446
 Qy 2497 --ANDEPKPLCSQVETLSDSE 2517
 Db 3447 VAAASESKPLLSKYDALSDSD 3469

RESULT 16
 Q7ZUE9 PRELIMINARY; PRT; 302 AA.
 AC Q7ZUE9
 DT 01-JUN-2003 (TremBrel. 24, Created)
 DT 01-JUN-2003 (TremBrel. 24, Last sequence update)
 DT 01-OCT-2003 (TremBrel. 25, Last annotation update)
 DE Similar to nuclear receptor co-repressor 1.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Body;
 RA Strausberg R.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC049302; AAH49302.1; --
 KW Receptor.
 SQ SEQUENCE 302 AA; 36126 MW; 8882177688970BC6 CRC64;

Query Match 6.3%; Score 827.5; DB 13; Length 302;
 Best Local Similarity 40.6%; Pred. No. 2.1e-32;
 Matches 181; Conservative 40; Mismatches 66; Indels 159; Gaps 6;
 Qy 11 TWRATEPRYPHSLSYVQIARTHTDVGLELQYHSHRSYASHLSPGSIQORRRPSLLS 70
 Db 12 SFSESGSRYPHSPVQVTFSTRHPQFVSVDYRNPLQD-----QORRRPSLLS 59
 Qy 71 EFGPNRQELHLRSHSYLPEL-----GKSEMEFIESKRPRLELLPDLRLPSLLAT 126
 Db 60 EPHPGTERPE-----RRHGYEQQFHAISAQOEQALESRRPIE----- 99
 Qy 127 GQPAGESEDLTKRSLTGKLEPVSPSPPHPTDPELELVPPRLSKEELIQNMORVDREITWV 186
 Db 100 -----NVS 102
 Qy 187 EQQISKLKKQOOLEBAAPPEKPVSPPTIESKHSLSVLQIYDENKKAEEAHRILLE 246
 Db 103 ETHEFSR----- 108
 Qy 247 GLGQVQLPLYNOPSDTRYHENIKINOAMRKKLILYFKRRNHARKQWKFCQRYDQLM 306
 Db 109 -----SQSDTKVTHDNIKINQVNRKLLILFFKERNHARKQWKFCQRYDQLM 157
 Qy 307 EALEKQVERIENPPRRAKESKVEYKEQFPEIRKORELQERNQSRVQSGLSMSAA 366
 Db 158 TEWEKKVERMENNPRRKAESRTREYERQFPEIRKOREQOQERFQ-RVGQRTGLSATIA 216
 Qy 367 RSEHVSIIIDGLSEQENLEKQROLAVIPMLYDADQRIKFTNNMGLMADPMKYKOR 426
 Db 217 RSEHISIIIDGLSEQENNEKQQLSVIPVMTDSEQRVKTNNMGLMDDPMKYKSR 276
 Qy 427 QVMNWSQEXETREKFMQHPKNGF 452
 Db 277 QFMVWTEHEKILGRSLCSIPRISG 302

RESULT 17
 Q86YY2

ID Q86YY2 PRELIMINARY; PRT; 453 AA.
 AC Q86YY2
 DT 01-JUN-2003 (TremBrel. 24, Created)
 DT 01-JUN-2003 (TremBrel. 24, Last sequence update)
 DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
 DE NCOR isoform d.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC YU L.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF303584; AAC32940.1; --
 SQ SEQUENCE 453 AA; 49417 MW; ADC3D56F36EADB20 CRC64;
 Query Match 4.9%; Score 650; DB 4; Length 453;
 Best Local Similarity 39.9%; Pred. No. 1.3e-23;
 Matches 181; Conservative 60; Mismatches 145; Indels 68; Gaps 22;
 Qy 2103 LPHLRPLPESQSSPLLOTAPG-----VKGHQVVTILAQHISEVITQDYTRHH--- 2151
 Db 28 LHHYRP---QGESPPQQLPPSSQAEQGVPRTHRLITLADHICQIITQDFARNQVSS 84
 Qy 2152 --PQQLSAPLPAPLYSFFGA--SCPVLDLRRPSPDLYLPPD-----HGAP-ARGSPH-- 2199
 Db 85 QTPQ---PPTSTFQNSPSALVSTPV---RTKTSNRYSPESQAQSVHHQPGSRVSPENL 138
 Qy 2200 ---SEGKRSPEPNKTSVLGGEDGTEPVSPPEGTEPGHRSRAVPLLYRDGEQTEPSR 2256
 Db 139 VDKSRGSRFGKSPERSHV---SSEPYEISPQ--VPVVHEKQDLSLLLSQGAEPAEQR 193
 Qy 2257 MGSKSPGNTSQPPAFPSKLTESNAMSVMKSKQEIKNKLTNRNPEFYNISOPGTEIFNM 2316
 Db 194 NDARSFOSISYLESPFKL-ENTSPVMVSKQEIFRKLNSGSGSDMAAQGTETIFNL 252
 Qy 2317 PAITGTGLMYRSQAOEHAHSTNMLEAIRKALCKYDQWEE-----SPPLSANAFNP 2370
 Db 253 PAVTTSVSSSRGSHSPADPAS-NLGLDIIRKALMGSDKVEDHGVVMSQPMGV---VP 308
 Qy 2371 LNASASLPAAMPITAAADGRSDHTLTSPGGG-GKAKVSGRPSRKAQSPAP--GLASGRDP 2427
 Db 309 GTANTVS-----VTSETRREEDGDPHSGVCKPKLISNSRSKSPKPGQYLGTERP 363
 Qy 2428 PSVSSVHSEGDGCMRRTPLTNRVWEDRPSAGSTFPFNPPLMELQAGVMAAPPPGLPAG 2487
 Db 364 SSVSSVHSEGDYHROTP--GWAWEDRPSSTGTQFPYPLTMRM---LSSTPTPTIACAP 418
 Qy 2488 SG-PLAGPH---HAWDEPKPLCSQVETLSDSE 2517
 Db 419 SAVNQAAPHQONRIWEREPAPLISAQVETLSDSD 452
 RESULT 18
 Q8BK32 PRELIMINARY; PRT; 291 AA.
 AC Q8BK32
 DT 01-MAR-2003 (TremBrel. 23, Created)
 DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
 DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
 DE Weakly similar to BB329D4.2.1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Body;
 RX MEDLINE=22354683; PubMed=1466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;

QY	935	NRLLSPRRSULLT--PFGDPRANASQKPLDLKQLKQRAAAIPIPIQVTKVHEPPREDAAPT	992
Db	6879	KRVWSPPPEVETVYVPEIPE-----KEPVELEKYEKV-----PP---TKPKDEEEDKGY	6927
QY	993	KPAPPAPPPPO-----NLQPESDAPOPQSSPRGSRSPAPPADKEAFAAEAQK-	1041
Db	6928	ERKPKDKPBEEDRKULGKGKURPEEGSEKLLKPKRKPSEKEAEKPOLKPIPKK	6987
QY	1042	-----LPGDPFCWTSGLPFVPPREVI-----KASHPADPDSAFSAYPPGHL	1084
Db	6988	PEBEKKDVTPKGGKP-----SKKIPDREPVELEPTEPEITLDKDKVPLEKPLPK	7041
QY	1085	PLGLH-DTARPVLP RP-----PITSNDPPLISSAKHPSVLERQIGAISQMSVQLHVPYSE	1139
Db	7042	PEPKESIEPEAPKLEEKSEBEPKUKGRPKPEKKEB-----	7083
QY	1140	HAKAPCVGVTGLPLMDPKK-----LAPFSGVKQEQLSPRGAGPPELSIGVPTAOEA	1192
Db	7084	-----PEVPSWRGKRLP--PKBEEKIEVLKPKPKKPEPKPKVKP---GKPYEPEI	7133
QY	1193	SVLRGTAL-----GSVFGSGITKGI PSTRPVPSDAITYRGSITHGTADLVLYKGTIT	1244
Db	7134	PEPEKTPLEPYTKDKEKVPDG-VTEPVK-----PED-----	7164
QY	1245	RIIGEDSRLDRGSDSLPKGHVIEYEGKHVLSYEGGMSVTQCSKEDGRSSGPPHET	1304
Db	7165	-----EEXP-----QPEEIKPKKERIKPKDDEEVETPSWGKRLPKPEEDKKEITLKPFK	7216
QY	1305	AAPKXTYDMMGRVGRAISSASIEGLMGRAP--PERH--SPHLLKEQHIRGSI-----TQ	1357
Db	7217	EKPKEPKSPKIKGCKYEP-----VIPEKPPLEPYEKPEREKVPQVQPKPEKE	7267
QY	1358	GIPRSYVEAQBDYL---RRRAKLLKRGTP-----PPPPSRDLTAYKQTALGPLKLKP	1409
Db	7268	KVPEQVPEPEKEISPKPEREKPQOLEKI PKRRKPPMLAAPIEKFSEIPKI-----	7320
QY	1410	AHEGLVATVKEAGRSIHEIPE-----ELRH--TPELPLAPRLKEGSIITQGLKXYDT	1461
Db	7321	-----TLKTKITQNVF-VPEBVTLETIELEHVEFTEV-IEPEVEKR---VMSPPPEVET	7368
QY	1462	GASTTGSKKHDRSLIGSPGRTFFPVHPLDMADARALERACVYESLKSRPGTASSGGS	1521
Db	7369	YVPEEIPKEPV--VLEKVEKYEPPTPKOEBEE---DKGYTERPKOKP-----	7413
QY	1522	IARGAPVIVPE-----LGKRPQSPLYTEDHGAFAGHLPRGSPVTWREPTLRLOEGLS	1575
Db	7414	-----EPEBDRKLGLGKGLRP-----EEEGEKKLKPPRRPKPSEKEAEKPOLKPIP	7462
QY	1576	SSKASQDRKLTSTPREIAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPL--APD	1633
Db	7463	KKKPEEEDKVTTPK-GKPSKKIPDREPVELSPFERTBPEI---LDKDKVLEKPLK	7517
QY	1634	PTSIPRIGIFLDAAYYIPRHIAPNPTVPHLYPPYLINGVPTDALENRTIINDYITSQ	1693
Db	7518	PKPEPEKPSIPEA---PKPLEEPSBEE---EPKLKEREKPKPEKEEBEV-----	7564
QY	1694	QMHNATATAMAQADMRLG--LSRPRESSIALNYAAGPBGIDL---SQVPHLPVLVPP-	1746
Db	7565	-----PSWRGRUPLPKEEB-----KEEIVLPFKKEKEPEPKPKPKY	7601
QY	1747	TPGTATAMDRLAYLPTAPQPFSSRHSSPLSP-----GGPHLTFKPTTSSSERERD	1799
Db	7602	KPGKP-----YEFIEIPEP-----EXTPLEPYTKDKEKVPDGVTEPVKPEDEEKP	7648
QY	1800	RD-----RER---DRDREKESILSTTVEHAPWRRPCTEQSSGSSSGGGSSSRP	1851
Db	7649	EEBIPKPKERIKDKDEEVE-----TPSWR-----	7673
QY	1852	ASHSHAHQHSPISFRTQDALQORPSVLHNTGMKGIIITAVEPSKPTVLKSTSTSSVPRAA	1911
Db	7674	-----GKRLPKEED-----KEEIT-LKPPKKEKPKPEKPSKIPGK	7710
QY	1912	TFFPATHCPLGGTLDGVYFTLMEPVLLPKEAP-----RVAPPEPRADTGHAFKAPP	1964

Db	7711	PYEP-----VPEPEKPLEYEPERKVPDQVPKEKEKV-----	7747
QY	1965	ARSGLEPASSPKGSEPRPLVPVPSGHATIIARTFAKNIAPHHASGPDPPAPPASADPHR-	2023
Db	7748	-----PQVPEPEKEISPKPEREEKPQPEK--IPGKRKRPPMLPAPITEEKFEI	7795
QY	2024	-----EKTQSKPFSIQEULRSLGVHG-----	2056
Db	7796	PKITLKTTQNVFVPEEVTLETIELHEVTEPVEIPEVEKRVWTPPPPYETVYVEIEPEK	7855
QY	2057	SPV-----SSPSLTHDKGLPKHEILDKSHLE-----	GELRPK 2089
Db	7856	EPVULEKYEKEPEPT-----KPKDEEEDKGYERKPKDPEPEDRKILKLGKGLAPE	7909
QY	2090	QGP-----VKLGGEAAHLPHLRPLPSQSSPPLQTA--PGVKGHORVVTLAQ	2137
Db	7910	EEGEKKLPPKRRPKSPKEAEKPOLKDIPKKKTEEEKVVTTPKPKKPSKEI----	P 7965
QY	2138	HISEVITQDYTRIHPQOL---SAPLPAPILYSFGASCPLDLRRRPPSDLYLPPDPHGAPA	2194
Db	7966	DREBFVELEPPFETPEILDOKVPLEKPLKPKPEPK-----EKSMEEVEVPKPEEKEP	8019
QY	2195	RGSPHSEGGKRSPEPNKTSVLGGEDGDI BFVSPPEGMTEPHGHSRAVYPLL---YRDGEOT	2252
Db	8020	K-----EEEPKPKVKRKEKDBEPEVSWRGRRLLPPEKEEKEEIVLKPFKKEKA	8071
QY	2253	EPSRMGSKSPGNTSQP--PAFFSKLTFESNSAMVSKKQKOEINKLINTHNNEPEYNISQ	2310
Db	8072	EPKPKRPTGKTYEPIPEPEKPLPEPYSPKPEKSPKPEEVLPKPSKPKKKEAEP-	8130
QY	2311	TEIFNMPAITGTGLMTYRSQAOVGHASTNMG-----EAITR--KALMGKYDQWEE	2359
Db	8131	-----KKFLVPVMAITEVEEAKVEETKATRVDIELTEVVTWKESVWKRVRNQPKDIEV	8186
QY	2360	SPPLSANAFNPLNASASLPAMPITADGRSDHTLTPGCGGKAKVSGRPS-----SR	2412
Db	8187	IPPSFSOFLEPI-----ITQTSHKSVFKCEVKGVPPEPKWYFNR	8226
QY	2413	KAKSPAG--LASGDRPPSVSSHSGD-----CNRRPTLTNRVWE---DRPS	2455
Db	8227	FEIIPSPNRILICYEDRVSTLTIVTTTDEDEGMYSCAEVNVAGKATTIANLVLEARKDKP-	8285
QY	2456	SAGSTP---PPYNPFLMR-----LQGVWASPPP	2481
Db	8286	-RGEAPKILEPLKPLKVKGHSKTVLETTVRGKPTP	8319
RESULT 20			
Q9W596 PRELIMINARY; PRT; 5412 AA.			
ID	Q9W596		
AC	Q9W596;		
DT	01-WAY-2000 (TREMBLrel. 13, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	CG3064 protein.		
GN	FUTSCH OR EG:49B4.1 OR CG3064.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Berkely;		
RX	MEDLINE=20196006; PubMed=107311132;		
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Mortman J.R., Vandell M.D., Zhang Q., Chen L.X.,		
RA	Branton R.C., Rogers Y.-H.C., Blazaj R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,		
RA	Abriel J.F., Agbayani A., An H.-J. Andrews-Frankoch C., Baldwin D.,		

QY 795 EPTFASATGAPTPPPAPPSPAPPPVPPVVEKEEETAAAPVVEGEOKPAPAEALVD 854
 Db 2418 QVSMTEITREDADQMPKPSOA-----ESRRESIAESIKASSPRDEKSPLASKEASR 2469
 QY 855 TGKAEPPVK-----SECTEABEGPAKGDAAEAETAEGALK-AEKKEGSGR 902
 Db 2470 PGSVAESIKYDLDRKQIIKODKSTEHRRSLEDKSAVTSEKSVSRPLSVASDHEAAVAL 2529
 QY 903 ATTAKSGAPOSD-----SSATCSADEVDEAE-----GDKNRLISP 940
 Db 2530 EDDAKGSLSPKDKRPGFAETVSPIEEATMBESKIEVVEKSSALSLQGGSGGKLQTD 2589
 QY 941 RPSLITPTGD-PRANASPO-----KPLDLKQKQAAAIPIPIQVTKVHEPPREDAAPT 992
 Db 2590 SSPVDVABEGDFSHAVASVSTVTLTKPAELAQ-----IGAATVSSPLDEALRT 2639
 QY 993 KPAP-----PAPPPONLOPESDAPQOPGSSPRGKSRSPAPPADKFAFAEAQKLP 1044
 Db 2640 PSAPPHIGRADSPAEACASEEIASODKSPQVLKESR-----PAWAEKDDAQLKSSVE 2694
 QY 1045 DPPCWTGSLPPVPPEVIK-----ASPHAPDPSAFSAPPGHPLPLGLHDTARPVL 1096
 Db 2695 D-----LRSPVASTEISRASAGETASSPIEAPKQDFA-----EFEQAEKAVL 2737
 QY 1097 P-----RPPTISNP-----PPLISSAKHPSVLEROIGATISQGSVQLHVPYSEHA 1141
 Db 2738 PLTIELKGNLPLTSSPVDVAHGDFPQTSTPTSSPTVASVQPAELSK-----VDIEKTA 2790
 QY 1142 KAPVGPV-----TWGLPLPMDPKKLPFGVQKQSLSPRGQAGPPPSLGVPTAOEA----- 1192
 Db 2791 SSPIDEAPKSLIGCABERPESPASAKDAESVEKSKDASRPSPVVESTKADTKGDIS 2850
 QY 1193 ----SVLRGTALGSPGGSITKGIPTSTRVPDSAITYRGSTHGTTPADVLYKGTITRIIG 1248
 Db 2851 PPSPEVLEG-----PKDVEKSKESRRPPSVSA-----SITGDSKDVSRPASVESVK 2899
 QY 1249 EDSFSLRDR-----GREDSLPGHVIYEGKGHVLSEYEGM--SVTQCSKE 1292
 Db 2900 DEHDKAESRRSIAKVESVIDBAGKSDSKSSQDSQKOEKSTLASKEASRESVVESSKD 2959
 QY 1293 DG-RSSSGP-----PHETAAP-----KRTYDME----- 1315
 Db 2960 DAEKSESRPESVIASGEPVPRESKSPDKDTSRPGSMVESVTADEKSEQOOSRRRESVAE 3019
 QY 1316 -----GRVGRATSSASIEGLMGRAPPPERHSPHILKEOHIRGISTQIGIPRSYVEA 1366
 Db 3020 SVKADTKDKGSKQASRSPSVDELKDDDEKOEESRRQSIGSHKAMSTGDESMDKADK 3079
 QY 1367 QEDYLARE--AKLLKEGTTPPPPP-----SRDLTEAVYKQALGPLKLPKPAHE 1412
 Db 3080 SKPSPRPESVAESIKHENTKDESPGLSRRDSVAESIKSDITKGEKSP--PSKEVSRPE 3137
 QY 1413 GLVATVK-EAGRSIHEIPRELR--HTPELPLAPRPLKEGSIQGT--PLKYDTGASTTG 1467
 Db 3138 SVVGSIKDEKAESRRRESVAESVPESSKDATSAP-PSKEHSRPESVLGLK-DEGDKTIS 3195
 QY 1468 SKK-----HDVSLIGSPGRTFPVPHPLDVMADARALERACVYESLSKSRPGTASSGGG 1521
 Db 3196 RRVSVADSIKDEKSLVQSASRPE-----SEAESLKDAAPSOETSRPESVTE---S 3245
 QY 1522 IARG-APVIVPELGPQSLTYEDHGCAFAGHLPRGSPVVTMBEPTPRLOEGLSSSK-- 1578
 Db 3246 VKDGKSPVASKEASRPASVAENAKSDADESKEQRP-----ESLPQKAGSIKDEKSP 3297
 QY 1579 -ASQD-----RKUTSTPREIAKSPHSTVPEHHPHLPSVYEHLLRGVSGVDLYRSHIPL 1630
 Db 3298 LASKDEAEKSKESRRRESVAEQFPLVSKVESRRSPASVAESV---KDEAEKSKESPLMSXE 3354
 QY 1631 AFDPTSIPIRGIPLDA-----AAAAYTLPHLIAPNPTPHLYPPYLIRGYPDATA 1678
 Db 3355 ASRPASVAGSVKDEAEKSKESRRRESVAESKSPPLPSKEASRPA----- 3396

QY 1679 LENRQTIINDYITSQOMHINTATATAQRAQADMLRGLSPRESSLALNYAAGPRGIIDLQVP 1738
 Db 3397 -----SVAESVKDEADKSKESRRRESGAESKSPLASKEA----- 3429
 QY 1739 HLPVLVPTPTGPTATMDRLAYLPTAQPFSSRRSSSPLSPGPGTHLTTPKTTTSSS--E 1795
 Db 3430 -----SRPASVAESI KDEAEKSKESRRRESVAESKSPPLPSKEASRPTISVAESVKDE 3479
 QY 1796 REEDRDRDRDREREKILTSTTTTVEHAPITWRPCTEQSSGSSSSGGGGSSRRPASHS 1855
 Db 3480 AEKSKESR-RDSVAESKSPLASKEASRPASVAESVQDEAEKSKESRRRESVAESKSPLASK 3538
 QY 1856 HAHQHSPTSPRTQDALQ--QRPSVLHNTGMGIITAVEPSKPTVL-----RSTSTS 1904
 Db 3539 EASRPASVAESIKDEAEKSKESRRRESVAESKSPLASKEASRPTISVAESVKDEAEKSKES 3598
 QY 1905 SPVRAATFPFPPATHCPLGTLGVVPTLMFVLLPKPEAPRVARPERPRADTGHAFKAP 1964
 Db 3599 SRDSVAESKSP-----LASKEASRPASVAESVQD--EAEKSKEE 3634
 QY 1965 ARSGLEPASSP---SKGSEPREPLVPPVSGHATIAKPAKNIAPHASP---DPPAPPASA 2018
 Db 3635 SRRESVAESKSPLASKEASRPASVAESVKDDEAEKSKESRRRESVAESKSPLASKEASRPASV 3694
 QY 2019 SDPHREKTOSKFSTQELRLSLGYHGSYSPEGVEPVSPVSSPSLTHDKGLPKHLEB-- 2076
 Db 3695 AESVKDEAEK---SKEESRRRESVAESKSPPLPSKEASRPTISVAESVKDEAEKSKESRESV 3751
 QY 2077 LDKSHLEGRLPKQPGPVKLGGEEAHLPHLRPLPESOPSSPPLQTPAGVKGHQVWVIA 2136
 Db 3752 AEKSKLASKASRPASVAESVKDEAEKSKESRRRESVAESKSPLAS--KEASRPASVA 3806
 QY 2137 QHISEVI--TQDYTRHHPOOLSAPLPAPLYSPFGASCPLDLRRPPSDLYLPPDHGAPA 2194
 Db 3807 ESKVDEAEKSKESRRRESVAESKSPLSK-----EASRPTISVAESVKDEADKSK 3854
 QY 2195 RGFSEHGGKSRPEPNKTSVLGGGEGDGI1BPVPPGEMTEPGHRSASVYLLYRDGEOQTP 2254
 Db 3855 EESRRRESGAESKSPASM-----EASRPTISVAESVK-----DETEK 3889
 QY 2255 SRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQKQKINKKLNHNNEPEYNIQOQTELP 2314
 Db 3890 SKESRRRESVTEKSP-LPSKEASRPTISVAESVKDEAEKSKESRRRESVAESKSPLASKES 3947
 QY 2315 NMPAITGTGLMTRSAQVQEHASTNMCLEAIIRKALMGKYDOWEESPPLSANAFN---PL 2371
 Db 3948 SRPASVA-----ESIKDEAGTKOESRRRESMPESGKAESIKGDQ 3986
 QY 2372 NASASLPAAMPITAAAGRSDHLLTSPGSGG-GKAKVSGRP-----SSRKAESKSPAGLASGD 2425
 Db 3987 SSLAKETSPPDSVVEKDETEKPEGSAIDKQVASRPESVAVSAKDEKSP-----HS 4041
 QY 2426 RPPSVS--SVHSEGDGNNRTPPLNTRV---WEDRPSAGSTFPFVNPPLIMRLQAGVWASPP 2480
 Db 4042 RPESVADKSPDASKEASRSLSAFTASSPIEEGPRSTADLSPLN-----LTGEA 4091
 QY 2481 PPGLPAGSGPLAGPHHAW-----DEEPKPLICSQ 2509
 Db 4092 KGKLTLPSSPIDVABEGDFLEKVAESSRPPAVLSK 4125

RESULT 21
 076891 PRELIMINARY; PRT; 5327 AA.
 ID 076891
 AC 076891;
 DT 01-NOV-1998 (TremBLrel. 08, Created)
 DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
 DE EG:49E4.1 protein.
 GN FUTSCH OR EG:49E4.1 OR CG3064.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

QY 1618 VSGVDLYRSHPL-----AFDPTSIPIRGIPLDAAAAYYLPRHAPNPTYPHYLPYPLIRG 1672
D 1619 VSGVDLYRSHPL-----AFDPTSIPIRGIPLDAAAAYYLPRHAPNPTYPHYLPYPLIRG 1672
Db 2990 VSVADSIKDEKSLVNSQAEARPESEASLK-DAEA-----PSQTSRPSVSTES 3037
QY 1673 YPDTAALNQTINDIYITSCQMHNTATACGADWLRLGSLPRESGLNALNVAAGRGII 1732
D 1674 YPDTAALNQTINDIYITSCQMHNTATACGADWLRLGSLPRESGLNALNVAAGRGII 1732
Db 3038 VKDCKSPVAGKEASRPASVAENAKDSADESKQRPESL-----PQVAGSIKDEKSLASK 3093
QY 1733 D-----LSQVPHLPVLVPTPGTPTATAMDRL-----AYLPTAPOPPSSHRSS 1775
D 3094 DEAKSKEESRRSVAEQFP-----LVSKEVSRPASVAESVKDEAKSKEESPLMSKEASR 3149
QY 1776 PLSPGCGTHLTKPTTTSSRRERDRDRDREREKSIITSTTVEHAPIWPGTTEQSS 1835
D 3150 PASVAG-----SVKDEAKSKEESR-RESVAESKPLPSKEASRPASVAESVKDEAD 3199
QY 1836 GSSGSSGGGSSRRPASHAHQHSIPSPRTDALQ--ORPSVLHNTGMKGIITVAEPS 1893
D 3200 KSEESRRSVAESKPLASKEASRPASVAESIKDEAKSKEESRRSVAESKPLPSKEAS 3259
QY 1894 KPTVL-----RSTSTSPVRPAATFPFATHCPLGGTLDGVYPTLMEPVLPLKEAPR 1944
D 3260 RPTSVAKSVKDEAKSKEESRDSVAEKSP-----LASKESR 3297
QY 1945 VARPERPRADTGHAFLAKPPARSGLEPASSP---SKGSEPRPLVPPVSGHATARTAKN 2001
D 3298 PASVAESVQD--EAKSKEESRRSVAESKPLAYKEASRPASVAESIKDEAKSKEESRR 3355
QY 2002 LAPHASPDPPAPPASADPHREKTQSPFISQIQLRLSLGHSYSPGVEPVPSPVSS 2061
D 3356 -----ESVAESKPLASKEASRPTSVAESVKDEAKSKEESRDSVAESKPLAS 3403
QY 2062 PSLTHDKGLPHL-BELDKSHLEG---ELRPKQPGVPLKGGEAHLPHL-----R 2107
D 3404 KEASRPASVAESVQDEAKSKEESRRSVAESKPLASKEASRPASVAESVKDDAEKSKEE 3463
QY 2108 PLPESOPSSPLLOTAPGVKGHVVTQAHTISEVI---TQYTRHPQOLGAPLPAFLYS 2165
D 3464 SRRESVAESKPLAS-----KEASRPASVAESVKDEAKSKEESRRSVAESKPLPSK--- 3515
QY 2166 FPGASCPLDLRPPSDLYLPPPHGAPARGSPHSEGGKRGPFEPNKTSLVGGGEGDIEPV 2225
D 3516 -----EASRPTS-----VAESVKDEAKSKEESRRSVAESKPLASKEA 3554
QY 2226 SPEGWTPECHRSAYVLLYRDGEOTSPRNGSKSPGNTOPPAFFKLTESNSAMVKS 2285
D 3555 SRPASVAE-----SVKDEAKSKEESRRSVAEKSP-----LASKESRPSASVAES 3600
QY 2286 KQOEINKLNTHRNE-----PEYNISQPGTEIFNMPAITGTGLMTYRQAOVQEHAST 2338
D 3601 VKDEAKSKEESRRSVAESKPLPSKEASRP-----TSVAESVKDEADK 3644
QY 2339 NMGLEAIIRKALMGKYDOWEESPLSANFNLNPLASASLPAMPITADGRSDHILTFPG 2398
D 3645 SK-----EESRRSVAESKPLPS-----ASMEASRPTSVAESVKDET----- 3679
QY 2399 GGGKAKVGRPSRKAASPAAGLASGRDPPSVS-SVHSEGDGNCR-----RPLTN 2447
D 3680 --EKSKEESRRSVAESKPLPS-KEASRPTSVAESVKDEAKSKEESRRSVAESKPLAS 3736
QY 2448 RVWEDRPPSAGST 2460
D 3737 KE-SSRPASVAES 3748

RESULT 22
Q81QA6 PRELIMINARY; PRT; 1966 AA.
AC Q81QA6;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CG6964-PA.

GN GUG OR CG6964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fierbin C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Hostin D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera R., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mitra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berkman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;


```
QY 2361 PPLSAN-----AFNPL-----NASASL----- 2377
Db 1812 PPIADTALQOLSEYARPHVAFSGVEQVYHHPMGPMYERLEETKNAQAAASQSRL 1871
QY 2378 -----PAAMPITTAADGRSDHTLTSPGGGKAKVSGRPSRKAASAPAGLASG 2424
Db 1872 DPHWMEYRRGIHPSQFPLVYANPAISQWERELGIPPHVGLDPGSHVMRMPQPEAGF 1931
QY 2425 DRPPSVSVHSEGCNRRRTPLTNRWEDRPSSAGSTPPFPYNPLIMRLOA 2473
Db 1932 QLPPNVG-----QYPRNMLIPREPHSDVLLRMSYADQLQYLOA 1970

RESULT 24
Q9BZSO ID Q9BZSO PRELIMINARY; PRT; 2406 AA.
AC Q9BZSO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Kappa B and V(D)J recombination signal sequences binding protein.
GN KRC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Thymocytes;
RX MEDLINE=21100880; PubMed=11161801;
RA Hicar M.D., Liu Y., Allen C.E., Wu L.C.;
RT "Structure of the Human Zinc Finger Protein HIVEP3: Molecular Cloning,
RT Expression, Exon-Intron Structure, and Comparison with Paralogous
RT Genes HIVEP1 and HIVEP2."
RL Genomics 71:89-100(2001).
DR EMBL; AF278765; AAK01082.1; -.
DR HSSP; P15822; 1BBO.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; P:transcription factor activity; TAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; TAS.
DR InterPro; IPR002034; AIPM/HcIt_synth.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNT_1; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 2406 AA; 259381 MW; 7F498F2BBF3AE93F CRC64;

Query Match 4.2%; Score 552.5; DB 4; Length 2406;
Best Local Similarity 19.8%; Pred. No. 5.2e-18;
Matches 553; Conservative 323; Mismatches 966; Indels 953; Gaps 136;

QY 3 GSTQLVAQTRATEPRYPHSLSPVQIARTHTDVGLEYQHGRDYASHLSPGS---II 59
Db 69 GSQKGTGQ--QKPKRPPIEASVHI-----SHVQPHLTPAFMSPGKEHLEGSTWQLV 122
QY 60 QPQRRRP--SLISEFPQG-NERSQELHLRPSHSLVLP-----ELGKS 98
Db 123 SPMRLGPGSLLIA---PGLHPSQ---LLPSHASIIPPEDLPGVKFVFPSPQVSLKPT 176
QY 99 EMEFTESKRPR-----LELLPDLRLPSLLA-----TGO-----PAGESDITKD--- 138
Db 177 EEAHKERKPKQPGKIYOYCSRPCAKFSVLQKHRSHTGERPPGCGFSFKTKSNLY 236
QY 139 -----RSLTGKLEPSPSPPTDPELELVPPRLSKEELIQNMDRVDRITWVEQOI 190
Db 237 KHRKSHAHRIKAGLASGGMGYPH-GLMERIPGEEFEEP-----TEGESTDSEET 288
QY 191 SKLKKKQOOLEEAAKPEKPEKVPFPPIES---KHSLSVQIYIDENRKAACAHLLEG 248
Db 289 SATSGHPAELS-----PRPKQPLSSGLYSGSHSSHERCSLSQSSTAQSLIEDPPFPV 342
```

```
QY 249 GPOVELPLYNQPSDTRQYHENIKINQAMKKLI-----LYFKRRNHARK 292
Db 343 ERSSEHPLSHKPEDTHTIKQLALRLSERKKVIDEQAFLSPGSKGSTGYFSRSAAEQ 402
QY 293 QWKQKFCQYDQIMEALEKKVERIENNRRAKESKVRREYKQFPEIKQRELQERMOS 352
Db 403 Q-----VSPNTNAK-----SYAEIIFGKC-----G 423
QY 353 RVCGSGSLMSAAR-----SEHEVS-----EIIDGLSEQENLEKQMLQAVIPP 397
Db 424 RIGQRTAMLTATSTQPLPLSTEDKESLVPLSVPRTOVIEHTKLITINE-----AVVDT 478
QY 398 MLYDADQQRKIFINNMGLMADPMKYKDRQVMNMWSEQEKETFREKPMQHPKNGFIASF 457
Db 479 SEIDSVKPRSSLSRRSSMESP-KSSLYREPLSSHSEKTKPEQSLLSLQHPPTAPPVPL 537
QY 458 LERKTV--AECVL-----YYLTKKNYKSLVRSY-----RRGKSQOQOQOQO 502
Db 538 LRSHMPSAACTISTPHHFRGSYFDDHI TDSEALSRSRSHVFTSHPRMLKQPQAIELPL 597
QY 503 QOQOQOQPMRSPSSQEEKDEKEKEAEKEEKEPEVENDKEDLLKEKTDGSDGDNDEKE 562
Db 598 GGEYSSEEEGPGSS-----KDTASKPSD-----EVE-PKESLTKKT----- 632
QY 563 AVASKGRKTANSQRRKG-----RITRSMANEANSEAITPQOSAEALASGMLNESSRWTE 617
Db 633 ---KGLKT-----KGVIECNICGARYKRDNYEAHKYCYSELQIAKPIASGTHTS 682
QY 618 EEMETAKGLLEHGRNWSAIRMVSGKTSQCKNFYNYKKRQNLDEILQOHLKWEKER 677
Db 683 PE---AEKSOIER-EPWSQ-----MMHYKLGTTLEL 709
QY 678 NARRKKKAPAAASERAAPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGEC 737
Db 710 TPLRKRKEKSLGDEE--PPAFESTKSPGS-----P 740
QY 738 GPATVNNSSDTEISPSPHTEAAKDTQNGPKPPATLGADGPPPGPTTP-----PRRT 789
Db 741 GPSDAARNLPLESTKSP-AEPPSKSV-----PSLEGPTGFQPRTPKPGSGESGKERR 792
QY 790 SR---APIETPASEATGA-----PTPPAP-----PS 814
Db 793 TSKEISVIQHTSFEXKSDSLEQSGLEGEDKPLAQFPSPPPAPHGRSAHSLQPLKVRQPN 852
QY 815 PSAPPPVVPKKEKEEETAAAPPVVEEGEEOQPPAAAEELAVDTGKAEPEPKSECTEE--AEE 872
Db 853 IQVPEILVTEEPDRPDTEPEPPPKPEK-----TEEFQWQPSQTLAQLPAEK 900
QY 873 GPAGKD---AEAATAEGALKAEKKEGGSGRATTAKSGAPQDSDSSATCSADEVDEA 929
Db 901 APPKKRLRLAEMAQAQSGSGSFESSVP---LSRSPQESNVSLSGSSRSASFERRDHGKA 957
QY 930 EGGDKNLLSPRP-----SLLTPTGDP-----RANASPKP-----LKLKOLKORAAAIPP 975
Db 958 EADPDSMDMRPKPLGTHTMLTVPSSHHPHAREMRRSASEQSPNVSHSAMHTTRSKSFYGS 1017
QY 976 IQVTKVHEPPREDAAATKPAAPPAP-----PPQNLQ-----PESD-----AP--QQ 1014
Db 1018 LSLT-----GPSAPAPVAPPAGEAPERRKCFVLRSPSLSRPPESELEVA PKGRQ 1067
QY 1015 PGSSPRGKSRPAPPADKEAFAAEQAQLPGDPPCWTSGLPFP-----VPPREVTKA----- 1065
Db 1068 ESEEPQSSSKPSAKSLSQISSAATSHGGGPGGKPGQDRPALGPTVPVTEALQVFFHP 1127
QY 1066 -----SPHAPP-SAFSVA-----PPGH-----PLPLGLHD 1090
Db 1128 VAQPTLHEKPYLPVPVSLFSFQHLVQHEPGQSPFFFTQAMSSLSLSPYSNPPLPPLSFQ 1187
QY 1091 TARPVLPRPPTISNP-----PPLISSAKHPSVLERQI---GAISQGM----- 1129
Db 1188 A--PPLPLQPTVLHPGLHLPQLMHPANIPFRQPPSFLPMPYPTTSALSSGFFLPQSQ 1245
```

```
Qy 1130 -----SVQHVYSEHAKAVGVNTWGLPLPMDPKKLAPFSGVKQB-QLSPRGQAGPP 1181
Db 1246 FALQPLGVESHLPQIKTSLAPATGSAG-----LSPQSYSSDIRLPP---VAPP 1293
Qy 1182 ESLGVPTAQEASVLRGTALGVPGSITKGIPI---STRVPSDSALTIRG-----SITHG 1232
Db 1294 ASSAPTSAPP-----LALPACPDPMVSLVFPVRVQTNMPSYGSAMYTTLSQILLVTSQG 1348
Qy 1233 TPADVLY-----KGTITRIIGED-----SPSRLDRGREDSLPKGH-----VIYEKK 1274
Db 1349 SSAITVALPKPEPSKG---TIVCGADVHEVGFSGSLSEESQSRAPPTPLVAVPTLPERK 1406
Qy 1275 GHVLSYEGGMVTOCKEDGRSSGPPHETAAPKRTYDM-----MEGRVGRAISASIEG 1329
Db 1407 GTSLSSESILSL-----EGSSSTAGSKRVLSPAGSLELTMETQOQKVEKEEASKADEK 1461
Qy 1330 LMGRAIPP-----ERHSPHLKEQHHRGSIITOGIPRSVVEAQEDYLBREAKLL 1378
Db 1462 L--ELVKFCSVVLSTEDGKPEKSHL-----GNQGG-----RELEML 1499
Qy 1379 K-----REGTTPPPPSRDLTEAYKTOALGPLKL---KPAHEGLVATVKEAGRSIHE 1427
Db 1500 SSLSDSPSDTKIEIPLPLPALSHGQAPGSEALKYQPSGKPHRGLTPL-----S 1550
Qy 1428 IPREELRHTPELP-LAP---RPLKEGSIQTGTPLKYDTGASTGSKKHDRSLIGSPORT 1483
Db 1551 VKKEDSKQEPDPLSLAPPSLPLSE---TSSRPKASQEG---TDSKK-----VLQ 1594
Qy 1484 PPPVH-----PLDVM-ADARALERA---CYEESLKSPGPTASSGSIARGA 1526
Db 1595 FPSLHTTTNVGWCVLNYIKPHNHIQHADRSSVYAGWCISLYNPLPGVSTKAAALSLLR-- 1652
Qy 1527 PVIIVELCKPROSLTYEDHGAPFAGHPLGRSPVTMREPTPRLOE---GSLSSSKASQDR 1583
Db 1553 -----SKQVSKETYMTATAP---HPEAGRLVPSSSRKPRMTVEVHLPSLVSEGGXDL 1702
Qy 1584 KLTSTPREIAKSPHSTVP---EHPHPITSPYEHLLRGVSGVDLYRSHIPLAP----- 1632
Db 1703 ARVEKEEERRGEPEADAPASQGEFARIKIFE-----GG---YKSNEYVVVRGRGK 1753
Qy 1633 -----DPTSIPRGIPLDAAAAYLPRHL-----APNPTYPHLYPLRYNGYDPT 1676
Db 1754 YVCBEGCIRCKPMSLKKHIRTHTDVRPYVCKGHCFAPKTKGNLTGHMKSKAHKKCOET 1813
Qy 1677 AALENRQTIINDYITSQOMHNTATAMAQRADMLRGLSPRESSALNVAAGPRGIIIDLSQ 1736
Db 1814 GVLE-----ELEABEGTS-----DDLQDSEGRGSEAV----- 1842
Qy 1737 VPHLPVLVPTPGTATAMDRLAYLPTAPQPFSSRRHSSPLSPGCPHTLTKPTTTSSSER 1796
Db 1843 -----BEHQFSDL-----EDSDS 1855
Qy 1797 ERDRDRDRDREREKSLTSTTT-----VEHAPIWRP-----GTEQSSG 1836
Db 1856 DSDLDEDEDEDEESQDELSRPSSEAPPPGPHALRADSSPLGPPQPDAPASGTATRG 1915
Qy 1837 SSGS-----SGGGGGSSSRPA-----SHSAHOHSPIISR----- 1866
Db 1916 SSVSEAEELTASSCMSMSQSPGLPWLGAPLGSVEKDTGSALYKVPSPRRPWSPSKEA 1975
Qy 1867 -TQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPV----- 1907
Db 1976 GSRPPLARKHSLTKNDS-----SPQRCSPAREPOASAPSPGLHVDPCRGMGPLPCG 2027
Qy 1908 RPAATFPFPAHCPLG-----GTLDGVYPTLMEPVLLPKEAP-----RVA 1946
Db 2028 SPRLQLSPLTLCPLGRELAPRAHVLSKLEGTDT---PG-----LPRYSPTRRMSPGOAE 2078
Qy 1947 RPERPRADTCHAFIAPKPPARSGLSPASSPSKGE-----PRPLVPPVS-CHATIARTP 1998
Db 2079 SPPR-----SAPGKVALAGPGSPAGEHGFGGLAPRVLPFPAPLPHKLLKSP 2128
Qy 1999 AKNLAPHASDPDPAPPASADPHREKTQSKPFSIQLELRSLGYHGSSSPGVEPVSP 2058
```

```
Db 2129 ETCASPMQKA--ESRSPSCSPQPAH--PLSSRPFS-----ALHDPFHGHILARTEENI----- 2176
Qy 2059 VSSPSLTHDKLPGKHLKHELDKSHLEGLRPKQGP-VKLGGEAAHLPHLRPLPESQPS 2117
Db 2177 -----FSH-----LPLHSULHTRA-----PCPLPIGG-----IQMVQARPOAH 2210
Qy 2118 PLLQATP-----GVKGHQRVVVTLAQHISEVITQDYTRHHPPQQLSAPLPAP---LYSPFG 2168
Db 2211 PTLPLPGPTAAWVSGFGGSDLTGARE-----AQERGRWSPTSSSASVSPVAKVSKFTL 2265
Qy 2169 ASCPVLDLRPPSLYLPDPDHGAPARGSPHSEGGKSPENKTSVLGGEDGIEPVSP 2228
Db 2266 SS--ELEGRDYPKERERTGGGPGRPDPWTGHGTGAPAEPTTHS-----PCTPP 2312
Qy 2229 EGMPTEPGHRSADVPLLYVRDGEQTEPSRMGSKSPGN---TSQPPAFFSK---LTESNSA 2281
Db 2313 DTLPRPQGR-----RAQSWSPRLSPANPEPSATPPLDRSSVGCCLAEA-SA 2363
Qy 2282 MVKSKQOEINKKLNT-----NRNEPEYNISQP 2309
Db 2364 RFPARTNLGSESTRQDSPKPSGSGEPRAHPHP 2398

RESULT 25
Q9NHX6 PRELIMINARY; PRT; 1966 AA.
ID Q9NHX6
AC Q9NHX6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE GRUNGE.
GN GUG OR CG6964.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Erkner A., Roue A., Core N., Angelats C., Vola C., Fasano L.,
RA Kerridge S.;
RT "Grunge is required for proximal and ventral leg development in
RL Drosophila."
RL Submitted (DSC-1999) to the EMBL/GenBank/DBAJ databases.
CC -I- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -I- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AF217844; AAF34752.1; -.
DR Flybase; FBgn0010825; Gug.
DR GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR000949; ELM2
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF03154; Atrophin-1; 2.
DR Pfam; PF01448; ELM2; 1.
DR Pfam; PF00249; myb_DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 1966 AA; 208033 MW; 96AF90E2082E770C CRC64;

Query Match 4.2%; Score 552; DB 5; Length 1966;
Best Local Similarity 20.0%; Pred. No. 4.3e-18;
Matches 490; Conservative 235; Mismatches 834; Indels 888; Gaps 111;

Qy 328 KVRYYEK-----QFPEIRKORELOERMQR-----VCQRSGLSMSAARSEHVESE 374
Db 16 QVNDVYAKLPDYNPISSFP-IDKETDEREESRWSPGVADGDLMLFLRAARSGMAAFQG 74
Qy 375 IIDGLSQQENI-----EKQMRQLAVIPMLYDAQORIKFINMGLMADPMKYVDKQVM 429
Db 75 MCDGLEDDGCUAASRDTPINALVDHDSGVDPKA-----LQMLVKPCKSGDKKK-- 126
Qy 430 NMWSEKETPREKFMQHPKNFGLI-ASFLEKRTVAECVLYYLTKNENYKSLVRRSYR 488
```


Qy 2361 PPLSAN-----AFNPL-----NASASL-----PAAMPI 2383
 Db 1812 PPYADTALRQLSEYARPHVAFRELEIKNAQAAASQSLRDLPHWMEYRRGIHPSQPL 1871
 Qy 2384 TAADGRSDHTLTSFGGKAKVGRPSRKAKSPAGLASGRPSV 2430
 Db 1872 YANPAISQMERERLGIPPPHVGLDPGSHVMRMPQPPEAGFQLPPNV 1918

RESULT 26
 Q8T9N4 PRELIMINARY; PRT; 1985 AA.

AC Q8T9N4
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Transcriptional corepressor Atro.
 GN GUG OR ATRO OR CG6964.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21652534; PubMed=11792320;
 RX Zhang S., Xu L., Lee J., Xu T.;
 RA "Drosophila atrophin homolog functions as a transcriptional
 RT corepressor in multiple developmental processes.";
 RL Cell 108:45-56(2002).
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -|- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
 DR EMBL: AF475087; AAL78679.1; --
 DR FlyBase; FBgn010825; GUG.
 DR GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
 DR InterPro; IPR002951; Atrophin.
 DR InterPro; IPR000949; ELM2.
 DR InterPro; IPR001005; Myb DNA binding.
 DR Pfam; PF03154; Atrophin-1; 2.
 DR Pfam; PF01448; ELM2; 1.
 DR Pfam; PF00249; myb DNA-binding; 1.
 DR SMART; SM00717; SANT; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 1985 AA; 210423 MW; 120A78F6C8FD67F CRC64;

Query Match 4.2%; Score 552; DB 5; Length 1985;
 Best Local Similarity 20.0%; Pred. No. 4.4e-18;
 Matches 505; Conservative 242; Mismatches 855; Indels 920; Gaps 117;

Qy 328 KVRVYEK-----OPPEIRKORELQERMSR-----VQGRGSLMSAARSEHEVSE 374
 Db 16 QVNDVYAKLPYNFISFP-IDKTDERELEESRWSGCVADGDLMLFLRAARMAAPQG 74
 Qy 375 IIDGLSEQENI-----EKQMRQLAVIPMLYDADQORIKFINMGLMADPMKYKDRQVM 429
 Db 75 MCDGLEDGGLAASRD DDTINALVDHSDYDPGKA-----LQALVKCPVSKGDKK-- 126
 Qy 430 NMWSEQEKETREKFMQHPKQFGLI-ASFLEKRTVAECVLYIYTKQENYKSLVRYSR 488
 Db 127 --WTEDETKKFIKGLRQFGKNFFRIHKDLLPHKDTPELVFYLWKKTPGANN--NRPHR 182
 Qy 489 RRGKSQQQQQQQQQQQQQQQMPRSSQBEKKEKEKEKEKEKEKEKEKEKEKEKEKEKE 548
 Db 183 RRRGSLRRNRVTRANNSNWTP-----PKKEDTPEPT----- 216
 Qy 549 KTDGTSGEDNDEKAVASKGRKTANSQGRKGRITRSMANESAEAITPQSSAELASME 608
 Db 217 -----ATTATAATAASE-----TASRSSPAVSKE----- 241
 Qy 609 LNESSRWTEEB-METAKKGLLEHGRNWSAIARMWGSKTVSOCKNFYNYKRONLDEILQ 667
 Db 242 --ENSSLTDDASECDSDSLTHKRDESP--SRM-----RTRN----- 275

Qy 668 QHKLMEKERNARRKKKAPAAASEEAFPPVVEDEMEASGVSGNEEMVEEAEALHAS 727
 Db 276 -----KQNNNSSTSGNNTAGNGGNATSISSGSTGGGAAGCNSSSKQDSANAV-AN 327
 Qy 728 GNEVPRG-----ECSPATVNNSSDTEIPSHTEAAKDTGONGPKPPATLADGPPPP 783
 Db 328 GKRPRGSETPDVSGGASVDS-----PKTPTTAKVAESSANKRK-----GKQDTP 372
 Qy 784 TPRRTSRAPLEPTPASEATCAPTPPAPPSP-----SAPPPVVPK-EKEEBETAAP 835
 Db 373 NKKRTTESNEPS-AHEENAIKEKRKRPDSFVSMNSDSRPSDVLDDGESNTTTT- 430
 Qy 836 PVEEGEQKPPAAEELAVDTGKABEPEVKSECTEAEAGPAKGDAAE-AEATAEGALKA 894
 Db 431 -----EQGSTKDSKE-TVSCKEEREMVNDLEAKAE--KAIRAEALAEKSDSAIKNM 481
 Qy 895 KKEGSGRATTAKSSGAPQSDSS-----ATCSA--DEVD 927
 Db 482 DEE-----TNIQAPSSADTSLVDGPNALPSPVAAPITMKVPTIATVEALNASVD 532
 Qy 928 EAEAGDKNRLLSRPSSLTPTGDPANASPOKPLDLKOLKORAAAIPIQVTKVHEPRE 987
 Db 533 RKEAIERMESCDSPMLKLATIKQEVSPQQQHMQQSQ-----QQMQ 578
 Qy 988 DAAPTK-PAPPAPPPQN-----LOPESDAPQPGSSPRGKSRSPAPPADKEAFAAEA 1039
 Db 579 QLAPVGIPQPPSCPPSVYIKPEMEDSMDATCNQNSNEPDLKVKIEIKNEDALKHSA 638
 Qy 1040 QKLPDPPCWTSGLPFPVPPREVIKASPHADPSAFSYPG--HPLPLGLHDTARPVL 1097
 Db 639 GGLPPSGPC-----APPSALHPL----- 656
 Qy 1098 RPPTISNPPPLISSAKHPSVLEROIGALSQMSVOLHVYPSE-HAKAPVGPVWCLPLPM 1156
 Db 657 -----SGAP--VESQEPHLQ-----HMPHQVTTQPPPGVYLIDG----- 690
 Qy 1157 DPKKLAPFGVKQQLSPRGQAGPPESGLGVPTAQEASVLRGTALGSGVGGSTTKGIPST- 1215
 Db 691 -----QLKYPSGGQVPP-----QPQLHSDAAGVSGA--PGAPITP 727
 Qy 1216 -RVPDSAITVRSITHTGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKK 1274
 Db 728 QKYPPEMEMKF-----APQDLKY-----PPPPPLD----- 752
 Qy 1275 GHVLSYEGMSVTQCSKEDGRSSSGPPHETAPKRTYDM-----MEGRVGRASSASIEG 1329
 Db 753 --ALKYSOEMQAAAA-----AAAAKYDMKYMMEQOQKYNVELSAH--- 793
 Qy 1330 LMGRAIPPER-----HSPHLKEQHIIIRGSIQTGIPRSYVEAQEDYLR 1372
 Db 794 -----QPPSKPGYQDSLKIPIKPGFGLPHNV-----GSP--LDAAHKY-- 831
 Qy 1373 REAKLLKREGTPPPPPSRDLTEAYKTQALGPLKLPKPAEGLVATVKEAGRSIHIPREE 1432
 Db 832 -----GPPPTSQESQQQPO-----PPAHQVPPGATPPGIA---MPKPH 868
 Qy 1433 LRHTPELPLARP-----LKEG---SITQGT--LKVD-----TGASTTCKKHVRS 1476
 Db 869 YQHDVQTPPLGRPEFTGLMLKYGDPLAAKYGPPQDLKYPMPVPSQAGPADVKYGGENL 928
 Qy 1477 IGSGRTPFPVHPLDVMADARALERACVEESLKRSPGTASSSGSGSIARGAPVIVPELKP 1536
 Db 929 IKSPYGFPPSPIDA-----SARSTPG--QDSQGSNSNSQPSMPPPOQ 972
 Qy 1537 RQSLTYTHDGAPFAGHLPRG-----SPVTMREPT----- 1566
 Db 973 FQSPHSPHMPSPAGGGLPGMHQPNLIHGPPPGAAGSGGPPPPPTSLHQPTPSAGP 1032
 Qy 1567 PRLOEGSLSSKASQDKLSTPREIAKSPH--STVPEHHHPH-ISPVEHLLRGVSGVDL 1623
 Db 1033 PSLOGLHLPQHQSQAASSIPPSSIGIPPTLSTMAPSHMHLPHAH-LOGL----- 1086


```
Qy 1765 -POPFSSRRHSSPLSPG-----PTHL-----TKPTTS 1792
Db 1234 HPGHLLSHSITAGLPGGGPIALLAGPGLGIPESALSRRTPSHLPHGHASSAPLTAH 1293
Qy 1793 SSERDRDRDRDREREKSLTSTTVEHAPWR--PGTEQSSGSSGSG-----CGGG 1846
Db 1294 SVA-----SMTSTSMSTLTSTVPSSAFSRASPSVQISSGGSPGSGVPGG 1341
Qy 1847 ---SSRPASHSHAH-----QSPISPERTQDALQORPSVLHNTGKGIITAV 1890
Db 1342 MPNSAAAAAHHRAASPASSVSSLSRQPLHPVQSPLSHPS-----SSALSAAAAV 1398
Qy 1891 -EPKPTVLRSTSSVPRPAATPPATHCPL-OGTLDGVYP-----TLMEPVLLP 1939
Db 1399 AERDRHALMRQS-----PHMTPPVSNASLMASPLSKWYAPQOGQGLGTSPPHLRP 1452
Qy 1940 KEAPRVAR-PRPRADTGHAPLAKPARSGLEPASPSSKGEPRPLVPPVSGHATARTP 1998
Db 1453 GASPPVIRHPOMPL-----PULIAPGGGIPQIGVHP 1484
Qy 1999 AKNLAPHASDPDPAPASADPHREKTQSKPFSIQELELSGLYHSSYSPEGVEVPSP 2058
Db 1485 GQSPYPH-----PULHPSVFSYSPHH-----PFN-----SPYGYAPYGP 1518
Qy 2059 VSSPSLTHDKGLPKHLELDKSHLEGLRKPQGPVKLG-EEAHLPHLRPL-----ES 2112
Db 1519 -----GFPAYMKP-----PPQQLDPAVAWAAHAGLQGPFPQOMRQD 1557
Qy 2113 QPSSSPLLQTAGVKGHQVVTLAQHISEVITQY-----TRHHPOOLSAPLPAP 2162
Db 1558 EQNAAAAAQAABEKOHAQAAAAAQAOKPQQOQPGGMPNKPPTPKTPQGGGMP-- 1615
Qy 2163 LYSFPGASCPLDLRRPSSDL-----YLPDPDHG--APARGSPHS-----EGKR 2205
Db 1616 ---PGMGPTPTGLPPGAYPGHMECYPOGPHGSPFPAPQDQPHGLKPTSIMDALRA 1671
Qy 2206 SPEPKTSVLGGGDBGIEPV-----SP-----PEGW-----TEFGHSRA 2240
Db 1672 HAHSANSAGMGHGHTEPLDIEPEPEIPSTHNIHPRGSPPEAKPDTECHRSOSA 1731
Qy 2241 VYPLYRDGEOTEFBSRMGSKSPGNTSOPPAFFSKLTESNMMVSKKQKQINKNLTNRN 2300
Db 1732 IFVRHIDRGDYNCTR-----TDLIFKPVADSKLA---RKREERDRKLAEKERE 1777
Qy 2301 EPEYNISOPGTEIFNMPAITGTGLMTYRSQAVQSHASTNMGLEALIRKALMGKYDWEES 2360
Db 1778 RROOQOQO-----QOQOQOQOAAAAQAQAQAQA-----ELK 1811
Qy 2361 PPLSAN-----AFNPL-----NASASL----- 2377
Db 1812 PPYADTPALRQLSEYARPHVAFSPVEQMVVYHHPMGPMYRELEETKNAQAAAAQSRL 1871
Qy 2378 -----PAAMPITADGRSDHLLTSPGGGKAVKVGSRPSRKAKSPAPGLASG 2424
Db 1872 DPHWMEYYRGTHPSQFPLYANPAISOEMERLERLGIPPPHHVGLDPGEHWMVMPQPEAGF 1931
Qy 2425 DRPPSV 2430
Db 1932 QLPPNV 1937

RESULT 28
O95875 PRELIMINARY; PRT; 2157 AA.
AC O95875;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE BAT2.
OS Homo sapiens (Human),
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
[1]
RN SEQUENCE FROM N.A.
RP Rowen L., Madan A., Qin S., Shaffer T., James R., Ratcliffe A.,
RA Abbasi N., Dickhoff R., Loretz C., Madan A., Dors M., Young J.,
RA Lasky S., Hood L.;
RT "Sequence of the human major histocompatibility complex class III
RT region.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF129756; AAD18086.1; -
SQ SEQUENCE 2157 AA; 228669 MW; 21B817F5B699B0DB CRC64;

Query Match 4.1%; Score 546.5; DB 4; Length 2157;
Best Local Similarity 20.4%; Pred. No. 8.9e-18;
Matches 514; Conservative 207; Mismatches 871; Indels 927; Gaps 114;

Qy 474 KKNENYKSLVRRSYYRRRGSKSQQQQQQQQQQ---QQQQQQQQMPRESSQEKDEKEKEAE 530
Db 70 KGNDPNVSLVKDGTGWASKQEQSDPKSDASTAQPPESQPLPAS-----QTPAS 119
Qy 531 KEEKKPEVNDKEDLLKEKTDGSDNDEKAVASKRKTANSQGRKGRITRSMANE- 589
Db 120 NQPKRPAAEPENTPLV-----PSGV---KSWAQAASVTHGAHGDGGRASSILSRSEEF 170
Qy 590 -----ANSEEAITPQOSAEILA-----SMEINSSRWTE-----EMETAKGLLSEHR 632
Db 171 PTLQAAGDQDKAAKAKERESABQSGPGSLRPQNSLTWRDGGGRGPDELE-GPDSKLHHGH 229
Qy 633 NWSA----- 636
Db 230 DPRGLQSPGPPFPYRGMPPFMYPIPPPPYPOGPPYPTDPGSRFRVAVGR 289
Qy 637 -----IARMVGSKTVSQCKNF--YFNKKRQNL-----EILQOHKLKMKERNARR 681
Db 290 GSGPPMELVEFVGRPSILKEDNLKFDQDQDNDGWAHAEEVDYTEKLKFSDEEDGRD 349
Qy 682 KKKK-----APAAASEEAAFPVVEDEMEWASGVSGNEEEMWEEAEALHASNEVPRG 734
Db 350 SDEEGAGCHRDOSASGEER--PP-----EADGKKGN-----SPNSPPT 387
Qy 735 ECSPATVNNSSDTE-----SIPSPHTEAAKDTGQNGPKPPATLGDGPPPGPTTP 786
Db 388 PXTAWAETSPPEPEPGPPAPKPPPLPPHRRGAGNWPAGPDYP-----DRGPPCKPAPPE 443
Qy 787 -----RTSRAPIE----- 801
Db 444 DEDEAWQRKQSSSEISLAVARRRRRREBERRMOEERRAAACAELKRLDEKFGADPKR 503
Qy 802 ATGAPTPPPAPPSAPSPVVPVVEKEEETAAAPPVVEEGEOKPPAAEELAVDTGKAEEP 861
Db 504 LKABPAAPPAAPSTPAPPVAVPKE-----LPAPPA-----PP-----P 536
Qy 862 VKSECTEEAEGPAKGDAAEAATAEALKAKEKGGSGRATTAKSGAPQDSDSATC 921
Db 537 ASAPTEKEPEEPAQAAPPAQSTTPP-----GVAAAPTIVSGGSTSTSSGSF 584
Qy 922 SADEVD-----EAEGGDKNRLLSPRSLTTPGDRANA-----SPQKPLD 962
Db 585 EASPVQLPSKEGPEPEEVPPTTPPVKVEKGDIGTRQPPSQGLGYPKYQSLP 644
Qy 963 LKQLKQRAAALPPIQVTKVHEPPREDAAPTKAPAPPPPPQNLQPEDAPQOQSSSPGK 1022
Db 645 PRFORQOQELLKQOQOQHQWQHQOQSGSAPPTFPVPPPPQVTL-GAVPAPOAP----- 696
Qy 1023 SRSPAPPADKEAFAAAEAQKLPDPPCWTSGLPFPVPPPREVIKASPHAPDPSAFYAPDGH 1082
Db 697 -----PPPPKALY-----PG-----ALGRPPPPPPMNF-----DPRWMMIPFYVD 731
Qy 1083 PLPLGLHDTARPLVLRPPTTISNPPLISSAKHPSVLEROIGAISQGMVQLHVPYSEHAK 1142
Db 732 P-----RLQGRFPLDFYPPGV-----HPSGLVPRERSDSGGSSE---PDRHAP 774
Qy 1143 APVCGVTNGLPLPMDPKLA-----PFGVKQEQ---LSPRCQAGFPFSLGYPTAQEASVL 1195
```

Db 775 AMLR--ERGTP-PVDP-KLAWGVDTATPAEPRPLTSPLRQAEDDDKGM----- 821

Qy 1196 RTGALGVPGSITKIGPSTRVPSDAITRGSITHGTADVLTKGTITRI-IGSDSPSR 1254

Db 822 -----RSETPPVPP-----PPYLASYGFPENGAGPP-----PISRPLEPPGPRP 863

Qy 1255 LDRGREDSLPKGHVIEYEGKGVHLSYEGGMSVTQCKEDGRSSGPPHETAAPKRTYDM 1314

Db 864 L-----FWPPG-----SDEVAKITQTPPKKEPPKEETAQLT 894

Qy 1315 EGRVGAISSAIEGLMGRAIPPERHSPHLLKQHHIRGSITQIGIPRSVVEAQEDYLRE 1374

Db 895 GPEAGR--KPARGVSGCGPPPPRRESRTETRWGPRFGSSRRGIPP----- 939

Qy 1375 AKLKAREGTPP-----PPPSRDLTEAVKTOALGLKLKPAHEGLVATVKEAGRSIH 1426

Db 940 -----EPCGAPRRAGPIKKPPPT-----KVELPP--KPLFQG-----D 973

Qy 1427 EIPRELRHTPELPLAPRLK-----EGSITQGTPLKYDTGASTTG----- 1467

Db 974 ETPK-----PKPDPLKITKGLGPKETPPNGNLSAPRLRRDYSYERVGPTSCR 1024

Qy 1468 -----SKXHDVRSLLGSPRTPPVHPLDVMADARALREACVEESLKSRCPTASSG 1519

Db 1025 GRGREYFARGRGFTYGGRR-----GARSREFRSYRE-FRGDDGRGGGTG 1071

Qy 1520 GSIARGAPVIVPELGPQRSPLTYEDHGAPFAGHLPRG--SPVTWRE-----PTPRLQ 1570

Db 1072 G-----PNHPP-----APRCRTASETSEGESEYEETPKRRQ 1103

Qy 1571 EGSLSSSKASQ-----DRKLTSRPE--IAKSPHSVTPEHHPIPISPYEHLLRGVSGVDL 1623

Db 1104 RGSETGSETSHDSLAPSDEAPTKEGTLTQVPLAPPPPGAPPPAPARFTARG----- 1157

Qy 1624 YRSHIPLAFDTSIPR-----GTPLDAAYLPL--RHLPN--PTYPHLYPP----- 1667

Db 1158 -----GRVTPRGVRRRGGRGPPVQPCGWSPPAKSLAPKPPPTGP--LPPSKEPLK 1210

Qy 1668 -YLIRGYDPTAALENRTIINDYITSQOMHNTATAMAQRADMLRGLSPRESSLAL--N 1723

Db 1211 EKLIPGLPLSVARGSGNSGVNMGEDGERPRRRHGRAQQQDK-----PPRFRLLKQREN 1266

Qy 1724 YAAGRGIIDLSQVPHLPVLVPTTGTATAMDRLAYLTPAQPPSSRHSSPSLPGGPT 1783

Db 1267 AARGSEKPSLT---LPASAP---APEEA---LTTVTVAPEARAAKSPDLNQNSD 1315

Qy 1784 HLTKPTTSS-----SERERDRDRE-----RD 1805

Db 1316 QANBEWETASBESDFTSERRGDKAEPVLLTPKAVTPGGGGGAVGIGSAMSRLDLSQ 1375

Qy 1806 RDREREKSILTSTTVEHAPITWRPQTEQSSGSSGSGGGG----- 1846

Db 1376 RAKOLSKESFSSQRPQMERQNRFPQCGKAGSSSSGSGGGGPGRTGPGRGDKRSWPS 1435

Qy 1847 ---SSSRPASHAHQHSPISPRTOALQORPSVLHNTGMKIITAVPESKPTVLRSTSTS 1904

Db 1436 PKNRSRPEERPPGLPLPPLPPSSSAVRLDQVIH-----SNPAGIQQALQ 1482

Qy 1905 SPVTPAATFPATHCPLGLTLDGVYPTLMPEVLLPKAPRVARPERPADTGHFLAKPP 1964

Db 1483 LSSRQGSVTPAGH-----PR-HKPGPPQAPQGPS--PRPP 1515

Qy 1965 AR-----SGLPASPSKSGEPRPLVPPVSG---HATARTPAK----- 2000

Db 1516 TRYEPQRVNSGL---SDPHEEPGPMVGVGTTPRDSAGVSPPPPKRERPPKPELLQ 1572

Qy 2001 --NLAPHAS-----PDPPAPPASAD-----PH-----REKTSKPFISQIELE--- 2037

Db 1573 EESLPPPHSGFLGSKPEGPGQAESRDTGTEALTPHWNRLHTATSRKSYRPSMPEPM 1632

Qy 2038 -----LRSLGTHGSSYSPEGVE-----PVSPVSSPSLTHDKGL-----PKH-- 2073

Db 1633 EPLSPFEDVAGTEMSQSDSGVDLSGDSQVSSGPGCSQRSSP-----DGLKGAAEGPPKRRP 1688

Query Match

4.1%; Score 543.5; DB 3; Length 2592;

Qy 2074 ---LEELDKSHLEGELRPKQPGPVKLGCEAAHPLHPLPESQSSSPSLQLQTAPGVKXH 2129

Db 1689 GSSPLNAVPCGPGSEPPRPPAPHDGRKELPREQLP-----PGPIGT 1735

Qy 2130 QRVVTLAQHISEVITQDVIIRHHPOOLSAPLPAPLYSFFGASCPVLDLRRPPSDL----- 2183

Db 1736 ER---SQH-----TDRGTEPGPIRPS-HRPGPPVQF-GTSDKSDLRLLVVGDSLKAKE 1784

Qy 2184 -----YLPDPDHGAPARG-----SPHSEGGKESPE-----PN 2210

Db 1785 LTASVTEAIPVSRDWELLPSAAASAEOSKNLDSGHCVPPESSSQRLYPEVIFYGSAGPS 1844

Qy 2211 KTSVLGGED-----GIEPVSPP-----PEGMTEPGHSRSV-YPLLYRDG 2249

Db 1845 SSQISGGAAMDQLHPNSGGFRPGTSLHPYRSQPLYLPPGPAPPSALLSGVALKQFLDF 1904

Qy 2250 EQTPSRMGSKSPGNTSQPPAFF-----SKUTESNAMYKSKQKQINKKLNTNRNE 2301

Db 1905 STMQATELGKLPAGGVLYPPPSFLYSPAFCSPLPDTSLQVRQ----- 1948

Qy 2302 PEYNISOPGTEIFNMPAITG--TGLMTYRSQAOEHASTNMGLEAI IRKALMGKYDQWEE 2359

Db 1949 ---DLPSP-SDFYSTPLQPGQSGFLPSGAPAQ-----MLLPWVDSQLPVVNFGLSPP 1998

Qy 2360 SPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRPSRSKAKSPAP 2419

Db 1999 APP---PAPPLSLPLVCPALQPPSLA-----VRPPPAPATRVLPSPA- 2038

Qy 2420 GLASGRPPSVSVHSE-----GDCNRRTPLTNRVWEDRPSSAGS 2459

Db 2039 -----RPFPAISLGRAELHPVELKPFQDYQKLSNLAGGPGSSRTPPTGRSFGSLNRLKA 2092

Qy 2460 TPFYPNPLI--MRLOAGVMASPPP-----PGLPAGSGP---LAGPHHAWDEEP 2502

Db 2093 TPSTYSGVFTQRVDLYQQASPPDALRWIPKWPWERTGLPPREGPSRRAEPEFGDKESP 2151

RESULT 29

Q9P3J0 PRELIMINARY; PRT; 2592 AA.

AC Q9P3J0;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Conserved hypothetical protein.

GN B7F21.040.

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.W., Mannhaupt G.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA German Neurospora genome project;

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 2 MYB-LIKE DOMAINS.

DR EMBL: AL389901; CAB97476.2; -

DR GO: GO:0005634; C:nucleus; IEA.

DR GO: GO:0003677; F:DNA binding; IEA.

DR InterPro: IPR001005; Myb DNA binding.

DR Pfam: PF00249; myb DNA-binding; 2.

DR SMART: SM00717; SANT; 2.

DR PROSITE: PS00090; MYB_3; 1.

KW Hypothetical protein; DNA-binding; Nuclear protein.

SQ SEQUENCE 2592 AA; 288030 MW; 22CC6601A164A2D8 CRC64;

Best Local Similarity 19.5%; Pred. No. 1.6e-17;
Matches 394; Conservative 256; Mismatches 735; Indels 635; Gaps 89;

Qy	60	QPORRRLLSEFOPG--NERSQELHURPESHVYLPGLGKSEM--EFIESKRRLLELLPD	115
Db	848	KPVQKPAQLDELPFRVSEPEVAPKPEPKVPSEPKKAPPELLPKLPEPEQIPR	907
Qy	116	PLLRSP-----LLATGQAGSDLTNRSLTGKLEFPVPPPPHTDPE	159
Db	908	POPOQPOPOQAQAQAPPLLTAEAPKSVTEPEREAEVQEAETQEQAP--TERQ	966
Qy	160	LELVP-----PRLSKEEL-----IQNMDRVREITWVEQQISKL	193
Db	967	LLAIPTVENEETSFLRPEQPKMEEMDTGSGGLPPLPTVEEHKVKVDQDMDVD	1020
Qy	194	KKQQOQLLEE---AAKPEPEKPVSPPIESKHSLSVLIYDENRKAEEAH	242
Db	1021	-----VDDEGRDALQP--PCKVPRLDVSDGSGESQFPRFPNQPTSDGASSRMDSDS	1072
Qy	243	-----RILEGLPQVELPLYNQPSDTRVHENIKINQAMRKK--LILY	283
Db	1073	EERTEDASVSGSVHEVRELSSTPTDELFPVN---VKPHQSKVKRLSEQSPGCAF	1128
Qy	284	F-----KRRNHARKQWKQFCQRYDQLMEALEKKVRIENPNRRRAKESKVRE	331
Db	1129	FMGNIRDMAETAKEHEA---KQVRYTNDAYLRF-----TLSDDP--TAVKSR---	1173
Qy	332	YYEQFPEIRKQRELQRMQS---RVGQSGLSMSAARSEHVESEII--DCLSE--QENLE	386
Db	1174	---NQFSQDKDKVSGKGHSGSDNHGKEGRRTTSRFSTELDYEAIAQESIREAQEKE	1230
Qy	387	KQMR-----OLAVIPMLY--DADQORTKFTNMGLMADPMKVYKDRQVMNW---	432
Db	1231	REERAQREKYTDKEAVIPEMIWTDKQHLFVDTSGLL--PLE-----KLVTQOAVP	1283
Qy	433	-----SEQKETFREKFMQHPKNFLIASFLERKTVAECVLYLYLTCKENYKSLVRSY	487
Db	1284	YHVNFTAEAEKFEKAYLEYKQWKGIAHELPNDFHFSVIQYYVAKKRELNLKRLKQP	1343
Qy	488	RRRKSQOQOQ-----OQOQOQOQOQOQPMRPSQEBKEKEKEAEKEEKEPEV	538
Db	1344	RRRKGRGKQYNALVSELGNPENETDNGENGNGRRQQPRRAAASFWGHEATPNA	1403
Qy	539	END-----KEDLLEKTDGSDNDEKEAASVASKRKTANSQGRKGR-----	582
Db	1404	DSDGATPSATGRRRAGTTTFPKNDSGAE---KPEGVKGGRRARQPKADKEPKVPKPAQ	1460
Qy	593	-----TRSMANEANSEEAITPOOSAE-----	603
Db	1461	AIAPTPPAASGKGTNARSRSNSTRQNPENMPKTPAELGARVQNPWFDPVPPGSMQPPLA	1520
Qy	604	-----LASMELNE-----	611
Db	1521	PAQOPTPLASPRAPPTLASTTISEVMAPPSLRPPPPAPASLPTFTGQSSGSPRIRTPQ	1580
Qy	612	---SRWTEEMETAKGILLEGRNWSALARMVGSKTVSQCNFYFNKYK--RQNLDELILQ	667
Db	1581	QASSVWSVSENDPSSLRSFGTDSALANHGKTKTQWVRYLYLRQKKEGQWEQIAT	1640
Qy	668	QHKLMKEKERNARRKKKAPAAASEEAAPFPVVEDEEMEAASVGSNGNEEMVEEBAALHAS	727
Db	1641	EADLK--KQGERPPPTPSAGPR-----KRYDVPSSSGHRPLAAAEAE-----	1684
Qy	728	QNEVPRGSCSPATVNS-----SDTESIPSPHTEAAKD--TGQNGPKPATIGAD--GRPPG	781
Db	1685	-PQAKSEAAFPANQPFQFQVPIQASVLHTLVQRSPVTNPAPLPPAPVAASASAPPS	1743
Qy	782	PPTPRRTSRAPIEPTPASEATGAPTPPAPPSAPPVVPVPEKEEETAAAPPEVERGE	841
Db	1744	QPTPSR-----PGPAVSQTMSPVPHPL---RQPTATVFVTEREGEPI---PV----	1784
Qy	842	EOKPPAAEELAVDTQKAEPEPKSECTEAEAGPKAGKDAEAAEATAEGALLKAEKKEGSG	901

RESULT 30
Q9HCL7
ID Q9HCL7 PRELIMINARY; prt; 2414 AA.
AC Q9HCL7;
DT 01-MAR-2001 (Tremblrel. 16, Created)
01-JUN-2003 (Tremblrel. 24, Last sequence update)

QY 922 SADEVYD-----EAGGDKNRLSPRESLITPTGDPANA-----SPQKPLD 962
DB 585 EASVPEQLPSKSGEPPEEVPPTTPPVKVEPKGDIGETRPQPSOGLGYPKYQKSLP 644
QY 963 LKQLKORAAAIPIPIQVTKVHEPREDDAAPTAPAPPPQNLQVLPESAPQPGSSPRGK 1022
DB 645 PRFQOQOEQLLKQOQHQHQOQSGSAPPVPPVPSFPQVTL-CAVPAPQAP----- 696
QY 1023 SRSPAPPADKFAFAEAQKLGDPDPCWTSGLPFPVPPREVIVKASPHADPDPSAFSYPFGH 1082
DB 697 -----PPPPKALY-----PG-----ALGRPPMPMPNMF-----DPRMMIPPYVD 731
QY 1083 PLPLGLHTARPVLPRPTTISNPPPLISSAKHPSVLERQIGAISQMSVQLHVPYSEHAK 1142
DB 732 P-----RLLOQRPLDLFPYPGV-----HPSGLVPRERSDSGSSSE-----PDRHAP 774
QY 1143 APVGPVTMGLPLMPDKKLA-----PFGVKQEQ-----LSPRGOAGPPESLGVPTAQEAQSVL 1195
DB 775 AMLR--ERGTP--PVDP--KLAWGVDTTATPAEPRLPLTSLRQAEDDDKGM----- 821
QY 1196 RTALGSVFGGSITKGIPTSTRVPSDAITYRGSITHGTADVLKGTITRI--IGEDSPSR 1254
DB 822 -----RSETPPVP-----PPYLASYPGFPENGAPG-----PLSRPFLPEPGRP 863
QY 1255 LDRGREDSLPKGHVYIEGKKHVLSEYEGMSVTCSEKEDGSSSGPPHETAAKTYDMM 1314
DB 864 L-----PWPPG-----SDEVAKITQTPPKKEPKKEBTAQLT 894
QY 1315 EGRVRAISSASIEGLMGRAIPPERHSPHLLKEQHHRGSIITOGITPRSYVEAQEDYLRE 1374
DB 895 GPEAGR--KPARGVSGGGGPPPPRRESRTETRMGPRFGSSRRGIPP----- 939
QY 1375 AKLLKREGTTP-----PPPPSRLTEAVKTOALGLKPKAHEGLVATVKEAGRSIH 1426
DB 940 -----EPCAPRRAGPIKKPPPT-----KVELLP--KPLEQG-----D 973
QY 1427 EIPRELKHTPLAPLPLK-----EGSITOGTPLKYDTGASTTG----- 1467
DB 974 ETPK-----PPKPDPLKITKGLGPKETPNGLSPAPRLRDRYSYVERVGTSCR 1024
QY 1468 -----SKKHVRSLSGRTFPVPHPLDMADARALERACVEESLKSRRPTASSG 1519
DB 1025 GRGGEYFARGRGFTGTGGR-----GARSREFRSYRE--FRGDDGRGGGTG 1071
QY 1520 GSIARGAPVIVPELCKRQSLTYEDHGAPFAGHLPRG--SPVTMRE-----PTPRLQ 1570
DB 1072 G-----PNHPP-----APRGTASETSEGESEYEELPKRRQ 1103
QY 1571 EGSLSSSKASQ-----DRKLTSTPRE--IAKSPHSTVPEHHHPITSPYEHLLRGVSGVDL 1623
DB 1104 RGSETGSETHESDLAPSDEAPTKEGTLTQVPLAPPPPGAPPAPARFTARG----- 1157
QY 1624 YRSHIPLAFDTSIPR-----GIPLDAAYLPL--RHLAPN--PTYPHLYPP----- 1667
DB 1158 -----GRVTPRGVPSRRGGRGPPPOVCGWSPPAKSLAPKPKPTGP--LPSKKEPLK 1210
QY 1668 --YLIRGYPDTAALNRQTIINDYITSOQHNTATAMARADMLRGLSPRESSLAL---N 1723
DB 1211 EKLIPGLPLSPVARGSGNSGVNMGEDGERPRRRHGRQAQODK-----PPRFRLLKQREN 1266
QY 1724 YAAGPRGIIDLSQVPHLVLVPPPTGTPTATA---MDRLAYLPTAQPPFSRRHSSPLSPG 1780
DB 1267 AARGSEG-----KPSLTLPASAPGPEEALTTVTAPAPRAAAKSPDLNSQ 1312
QY 1781 GPTHLTKEPTTSS-----SRERDRRE----- 1803
DB 1313 NSDQANEWEATASSSDFTSRRGRKEAPPVPLLTPKAVGTPGGGGGAVPGISAMSRGD 1372
QY 1804 -RDRDREREKILSTTTTVEHAPITWRPCTEGSSSGSSGGGGG----- 1846
DB 1373 LSQRAKLSKRSFSQRCMGRQNRPPGPGKAGSSSGSSGGGGGPGGRTGPGRGDKRS 1432
QY 1847 -----SSSRPASHGHAHQSPISPRTDALQORFSLVHNTGMKGIIITAVEPSKPTVLRS 1901

DB 1433 WPSKNRSRPEERPPGLPLPPPPSSSAVRLDQVIH-----SNPAGIOQA 1479
QY 1902 STSSPVRAATFPATHCPLGTLGDVYPTLMEPVLLPKEAPRVARPRPADTCHAFLA 1961
DB 1480 LAQLSSRQGSVTAFCGH-----PR-HKPGPPQAPQSPS--P 1512
QY 1962 KPPAR-----SGLEPASSKSGSEPRPLVPPVSG-----HATTIARTPAK----- 2000
DB 1513 RPPTRYEPQRVNSGL---SSDPHFEEPGVMRGVGTGTRDSAGVSFPFPKGRERPRKPE 1569
QY 2001 -----NLAPHAS-----PDPPAPPASAD-----PH-----REXTQSKPFIQBLE 2037
DB 1570 LLOBESLPPPHSSGFLGSKPEGCPQAEESRDTGTEALTPHWNRLHTATSRKSYRPSME 1629
QY 2038 -----LRSIGYHGSYSPEGVE-----PVSVPSSPSLTHDKGL-----PK 2072
DB 1630 PWMEPLFPFEDVAGTENSQSDSDVLSGDSQVSGFCSQORSP-----DGLKGAAGEGPPK 1685
QY 2073 H-----LEELDKSHLEGELRPQPGPVKLGEAAHPLRLPLPESQSSPSLLQTAFCV 2126
DB 1686 RPKGSSPLNAVPCGPGSEPPRPPAPHDGRKELPREQLP-----FGP 1732
QY 2127 KGHQVVVTLAQHISEVITQDITRHHPOQLSAPLPAFLYSFPGASCPLVLLRRPPSDL--- 2183
DB 1733 IGTER-----SQR-----TDRGTGEPGIRPS--HRPGPPVQF--GTSDDKSDRLVVGDSLKA 1781
QY 2184 -----YLPDPDHGAPARG-----SPHSEGGKSP----- 2208
DB 1782 EKELTASVTEALPVSRDWELLPSAAABQSKNLDGHCVPFESSQRLYPVEFYGSA 1841
QY 2209 -PNKTSVLGGED-----GIEPVSP-----PEGMTEPGHRSASV--YPLLY 2246
DB 1842 GPSSSOISGGAMDQLHPNSGFRPGTSLHLYRSQPLVLPFGAPPALLSGVALKGQF 1901
QY 2247 RDGQOTPSRMGSKSPGNTSOPPAFF-----SKLTESAMVSKKQKQINKKLINTHN 2298
DB 1902 LDFSTQATELGKLPAGGVLYPPPSFLYSAPAFCSPLDPTSLQVRQ----- 1948
QY 2299 RNEPEYNISQGTREINMPAITG--TGLMTYRSQAVQEHASTNMGLEAIIRKALMGKYDQ 2356
DB 1949 -----DLPS--SDFYSTPLQPGQSGFLPSCGAPAQ-----MLLPVDSQLPVVNFSG 1995
QY 2357 WEESPLSANAPNLINASASLPAAMPITAAADGRSDHTLTSPGGGKAKVSGRPSRKAAS 2416
DB 1996 LPPAPP---PAPPPLSLLPVGALQPPSLA-----VRPPAPATKVLPS 2036
QY 2417 PAPGLASDRPPSVSSVHSE-----GDCNRRTPLTNRWEDRPS 2456
DB 2037 PA-----RPPASLGRAELHPVELKPFQDYQKLSSNLGGGSSRTPPTGRSFGSLNSR 2089
QY 2457 AGSTPPFVYNLI--MRLQAGVMAASP-----PPPLGAPAGGLAGPHHAWDEEP 2502
DB 2090 LKATPTSYGVFRTORVDLYQOASPPDALRWIPKWNERTGPPREGPSRR--ABEP 2143

RESULT 32

Q9UQ35 PRELIMINARY; PRT; 2752 AA.
ID Q9UQ35
AC Q9UQ35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RNA binding protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ohtaki S., Umeki K., Sawada Y.;
RT "Homo sapiens mRNA for RNA binding protein, complete cds."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

Query Match	3.9%;	Score 521;	DB 11;	Length 2607;	
Best Local Similarity	19.1%;	Pred. No. 1.9e-16;			
Matches 550;	Conservative 325;	Mismatches 1091;	Indels 920;	Gaps 123;	
Qy	44	HHSRDYASHLSPG--SIIOQORRRPSSLSEFOP--GNERSQELHLR-----PESHYLYPEL	95		
Db	206	HQBGDAPSVFPGATNIQQPSSPAPSTKQSSSPYEDKDKKEKSAVPPSPSPERSSTGP	265		
Qy	96	GKSEWFIESKRPRLELLDPDLL-----RSPILLATQOPAGSDDL--TKDRSLT-GKLE	146		
Db	266	P-----APTELLVEQHDDSPRLAAI--PSQFPVPNPSSEASTRGCS	307		
Qy	147	PVSPSPPTDPELELVPRLSKEELIQNMORVDREITVMVEQQISKLKKQKQOQLEEEAAK	206		
Db	308	PKSPEKPPQSTSS-ESCPP-----SPQPTKVSRRHASSPESLKP	345		
Qy	207	PPFP--EKPVPFPPIESKHSRLVLIIDYENRKKABAAHRIIEGLQPOVELFLYNQPSDTR	264		
Db	346	TPAFGRREISSSP-TSKNRSHGRAKDKSHST-PSHRAGRSRSPATKRGSRSRSTPTK	403		
Qy	265	OYHENIKINOAMRKKILIFYKRRNHARKQKQFCQRYDQLMEALEKKKVERIENPRRA	324		
Db	404	RHGRSRSPQ-----WRRSRSAQRWKG-----SRSPQRRG	433		
Qy	325	KESKVREYERQFPPIRKQRELQERMOSRVQQRGSLGMSAARSEHEVSEIIDGLSEQEN	384		
Db	434	RSR-----SPQPCWRSRNTQRRGRSARRGRSHSRPATR-----GRSRRT	478		
Qy	385	LEKQMRQLAIVPMLYDADQOQRIKFINNGLMADPMKVIKDRQVNMWSEQEKTFREKF	444		
Db	479	PARRGRS-----RSRTPARRRSRRTPARRRSR	506		
Qy	445	MQHPKNEGLIASFLERKTVASCVLYYLTKKNENKSLVRBSYRRGKSQOQOQOQOQOQ	504		
Db	507	SRTPARRRGRS-----RTP-----TRRSRTRSPVRR--RSRGRSQARRSGRGRSR	551		
Qy	505	QOQOQOQPMRSPQEBQKEKEAEKEEKEPEVENDKEDLLKEKTDGTDGNDKEAV	564		
Db	552	TPARRSGRSRTPARRGRSRTPARRSAR-----SRSRTP	588		
Qy	565	ASKGRKTANSQGRKGRITRMANEANESEAITPQQAELASMEINLESSRWTEEMETAK	624		
Db	589	ARRGRSRRTPARRRSR--SRSLVRGRSHSRTPQRRGRSGSSSRKKNKRTSQ-----	640		
Qy	625	KGLLEHGRNWSAIARMVGSKTVSQCKNIFYNYKKRONLDEILOQHLKWEKERNARR---	681		
Db	641	-----RRGRSNSPPEMKKSHVSSRRSRSLSS-----PRSKAKSLRRSL	679		
Qy	682	-----KKKAPAAAEEAAPPVVEDEEAEAGVSGNEEEMVEAEALHASGNEVPRG	734		
Db	680	GSSPCPKQKSTPTTTRSRSGSPPKQKSKTTPQSRSNSPQPKV-----SGTP-PR-	730		
Qy	735	ECSGPATVNNSSDTESTPSP-----HTEAAKDTQNGPKPPATLGADG-----PP--	779		
Db	731	-----PGSVTNMQADECTATPQQRSHSSESP-GEVKSRTPRQSCSGSSPRVKSSTPPRQ	785		
Qy	780	-----PGP-----PTPRTSRAP-----IETPASEA-----	802		
Db	786	SPSRSSSPQPKVTVISPRGRSHSSSSPSPSRVTSRTPQKRSRSVSPCKPVDSRLRHSR	845		
Qy	803	-----TGAP-----TPPPAPSPSAPPVVPVVKKEKEETAAA-	834		
Db	846	SRSSSPDSKMLGTPLRHSGSTSYPYKSMLTQTPDQNLGSKSP--CPQKSRDFTGSSG	903		
Qy	835	-----PPV-----EGEEQKPPAAEELAVDTGKAEEPVKSECTEEAE	871		
Db	904	SFHLCPGVTSSIVPGESCFASFVQKGHTQTPDTSPPVMTQVESPLLOSKSQSTP	963		
Qy	872	EGPAKGDAEAAAEATAEGALKAEEKGSGRAITAKSSGAPQDS-----DSSA	919		
Db	964	KGSLSRSSSPVELTARSVPVKQDKSEISTD--PKLKGMSPEQSKTKPDSSIYPLVDSKS	1021		
Qy	920	TCSADEVDEAFGGDKNRLLS-----PRSLTTPGDPDRANASPOKPLDLKQKORAA	971		
Db	1022	FLVQSRLEPSELKELGLIQEDVASSCIPRDK-FSPQDRPESSTVLKVTVRVLKERSG	1080		
Qy	972	AIPPIQTVKHEPPREDAATKP-----APPAPPPQNLQOPESDAPQOPG--	1016		
Db	1081	AGSP-----PKGRDQKSLLPNSQDELMEVKEBQPLSQVLPSLSPEH--KEMPGSN	1130		
Qy	1017	--SSPRGKSRSPADPADEAFABAQAQLPGDP--PCWTSGLPFPVPPEVIKASPH-AP	1070		
Db	1131	IESSEPEVEER-PAVLSALDQSQSPSKAAGTFAVASCW-----SGPQVSP	1174		
Qy	1071	DPSAFSAPPGHPLPLGLHDTARVLPRLPPTI-----SNPPLISSAKHSPSLEROI	1122		
Db	1175	EKELSHSP-----PRENSFESSLEFKNSGPVSEVNTGFSPEVKEELN	1217		
Qy	1123	GAISQGMVQLHVPYSEHAK-----APVGP-VTMGLPLMPDKKLAPFSGVKQEQOLSFR	1175		
Db	1218	GSFLNQTEADPSVDMKQSKSRSSRSELSEPVVEKVGLFSSQKVSSPVLETVOORTFSR	1277		
Qy	1176	--CQAGPPESLGVPTAQEAASVLRGTAIGSVPGGSIITKGIPSTRVPSDSAITYRGSITHG	1232		
Db	1278	ERSSASAPELKDGLPRTPS-----RRSRGSGSPGLRDGSGTFS-----RHSLSGS	1322		
Qy	1233	TPADVLYKGTITRIIGEDSPSRLDRGR--DSLPGKHVIYEGKKGHVLSYEGGMSVTOCS	1290		
Db	1323	SPG-----MKDTPQTPSGRSECDSPPEPKALPQTPRAR-----SHSPSS	1362		
Qy	1291	KEDGRSSGPPHETAAPKRYDMM-----GRVGRAISSASIEGLMGRATIPPER---H	1340		
Db	1363	PERNNKSVTPQRESGSESSVEQKNLARTSPQGRSRSGSQELDG-KPSASPQERSESDS	1421		
Qy	1341	SPHH-----LKEOHIRGSIITQIPRSVYEAQEDYLRRKALLKREGTPPPPPPSRDL	1393		
Db	1422	SPDKSKTRTPLRORSHSGS-----PEVDSKSRHSPLSRSGSGSPPEMKDK---	1467		
Qy	1394	TEAYKTOALGPLKLKPAHEGLVATVKEAGRSIHIPREELRHTTPELPLAPRLKEGSTIQ	1453		
Db	1468	-----PRVLQRAQSGTSSPE-----HKIP-----APRALPHS-RS	1498		
Qy	1454	GTPLKYDTGASTTQSKKHVDVRLSLGSPCRTPPPVHPLDVMADARALEACVYESLSKR-P	1512		
Db	1499	GSSSK-ERGPSPEGSSSESSSEPEHAPKSR-----ARRGRSSIEPTKSRTP	1545		
Qy	1513	GTASSSGSITARGAPVIVPEL-GKPROSPLTYEDHGAPFAGHLPRGSVPTWREPTPLQE	1571		
Db	1546	PRRSRSS-----PELTRKARVSRRSRSASSP-----EIRSRTPRRRS	1587		
Qy	1572	GSLS-----SKAQDRKLTSTPREIAKSPHSTVPBHHPHPI-----	1608		
Db	1588	PSVSSPEPTKESRRRRRSVSSPRTKTSRRGRSPSPKPRGLQRSRSRREKTRTTRR	1647		
Qy	1609	-----SPYELLRGVGVLYRSHIPLAFDPTIP-----RG---I	1641		
Db	1648	RDRSGSQSTRRRQRSRSRVTRRRRGSG--YHSRSTPQRESSRSTRRRGRSRT	1704		
Qy	1642	PLDAAAAYLPRHLAPNTPVPHLYPPVILRGVPTDALENQTTIINDVITTSQOMHNAT	1701		
Db	1705	PLTS-----RKGRSRSTPAPWKRSRSRASP---ATHRRSRITPLISRRSRSTSP	1754		
Qy	1702	AMQADMLRGLSPRESSLAINYAAGPRGIIDLSQVPHLPVLVPPPTGCTATAMDRLAYL	1761		
Db	1755	VSRRS---RSVNRRRSRSRASPVRRR---SRSRTP-----PVTRRRSRST	1796		
Qy	1762	PTAPQPFSSRHSSPLSPGGPHTLTKPTTTSSSRERDRDRDRDREREKSIITSTTV	1821		
Db	1797	PTRRR---SRSTPVTTRRRSRSTPPTV-----RRSRSRSTPVTTRRRSRSTPVT	1848		
Qy	1822	EHAPIWPGTEQSGSGSGSGSGSGSSR--PASHSHAHQHSPISTQD-----ALQO	1873		
Db	1849	RSRSTPVTTRRRSRSTPVTTRRRSRSTPPTFATRRSRSTPPLPRKRSRSPALAIR	1908		
Qy	1874	RPSVLHNTGMKGIIITAVEPSKPTV-LRSTSTSSPVPAAATPPATHCPLGGLDGV----	1928		

Db	10765	KVPEPKVPEEKIRISITKRE---KEQVTEPAKVFPMKPK-----RVVAEEKVPVRKE	10816
Qy	898	GGSGRAITAKSSGAPQSDSSATCSADE---VDEAGGDKNLLSPRPSLLT-----	946
Db	10817	---VAPVRVPEPKLEPEEVAPEEVEVTHVEEYLVIEEEXYIIEEDEFITEEVEVPV	10872
Qy	947	-PTGDPRANASP---OKPLDLKQLKORAAAIPIQVTKVHEPPREDAAFTKAPAPPPP	1001
Db	10873	IPVKVPEVPRKPVPEEKVPVPPKKE---APPAKVPEPKPEEKVPVLIPIKKGKPPP	10928
Qy	1002	PONLOPESDAPQCGSGSPRGKRSRAP-PADKEFAFAEAQKLCDDPPCWTSGLPFVPPR	1060
Db	10929	---AKVPEVP-KRPVPEKVPVPEVPPKVEAPPAKVPEVPPK-----VPEK	10970
Qy	1061	EVIKASHAPDPSAFSAPPCH--PLPLGLHDTARPVLPRPPTISNPPPLISSAKHPSVL	1118
Db	10971	KVPVPAPKVE-----APPAKVPEVPPKLIPEEKKTTPVKVEAPPKVPKREP---	11021
Qy	1119	ERQIGALSQMSVOLHVPYSEHA--KAPVPGVTWGLP-----LPMDPKKLAPSGV--KQ	11169
Db	11022	-----VPPVVALQEEEVLFEEIIVPEEEVLPPEEEVLPPEEEVLPPEEEVLPPEE	11071
Qy	1170	SOLSPRGOAGPPESLGVTAQEAQSVLRTALGSGVPGGSITKGI-PSTRVPS-----	1223
Db	11072	EEIPPEEEVPEEEYVPEEE-----FVPEEEVLPPEVPPKVPVPAVPEIKKV	11121
Qy	1224	TVRSGI---THGTADVLXGTITRIIGEDSPGRDLRGREDSLFKGHVIEYEGKKGHVLS	1279
Db	11122	TEKKVVPKKEAPKVV-----PEVPKVEKR-----IILPKEEEVLP	11161
Qy	1280	YEGGMSVTCCKEDGRSGSPHETAAPKRTYDMMGEVGRGAISSASIEGLMGRAIIPER	1339
Db	11162	VE---VTEPEEPISEEEIPEEPPSIEEVEEAVPRVPEVIKAVEA--PTFVPKKV	11215
Qy	1340	HSPHLKQEHRTGSIQTGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKT	1399
Db	11216	EAP-----PAKSKKIPBEKVPVPPVQ--KKEAPPAKVPEVPKKVPEKVLVP--KK	11262
Qy	1400	QALGPLKLPKPAHEGLVA-----TVKE-----AGRSIHEIPREELRHTPELPLAPRL	1446
Db	11263	EAVPPAKGTVLEEKVSVAFRQEVVVVKERLEVEAEVEEIEPEEEFHEVE-----	11317
Qy	1447	KGSIQTGTPKLYDTCAGTTGSKKHIDVRSLSGPGRTFPPVHPLD-----VMADAPALERA	1502
Db	11318	ERGEF-----HEVEEFIKLEQHRVEEHRVKEVHRVIEVFEEAVE	11358
Qy	1503	CYEESLKSRRGTASSSGGSIARGAP--VIVPELGPQPSQ-----LTVEHGAFFAGHL	1554
Db	11359	VPEKP-KAPPKGPEISEKLIIPPKPTTKVVRKEPPAKVPEPKVIIVVEEKVRVP--EE	11414
Qy	1555	PRGSPVMTREPTPRLOEGLSSSKASQDKLTLSTPREIAKSPHSTVPEHHPIPSYEH	1614
Db	11415	PRVPTKVPDLVPKEVPEKKVPVPPAKKPEAPPKVPAPKEVVPKKV-PVPPPK--	11471
Qy	1615	LRQVSGVDLYRSHIPLAFDPTSTPRGIPLDAAAYVLPRLHAPNTPYHLYPYLIRGVP	1674
Db	11472	-----KPEVP-----PTKVPE-VPKAAVPEKKVPRAPIPKPESP--PPEVPEA-P	11512
Qy	1675	DTAALENROTIINDYITSQQMHNTATATAQRAQADMLRGLSPRESSIA-LNYAAGRPGIID	1733
Db	11513	KEVVPEKK-----VPAAPPKKPEVTPVKVPEAPKEVVP	11545
Qy	1734	LSQVPHLVLPVPTPGTATAMORLAVLPTAPOPFSRRHSSSPSLGCGTHLTKTPTTSS	1793
Db	11546	EKKVP-----VPP-PKKPEVPTTKVPEVKVAVPEKKVPEAIPPKPESP-----	11595
Qy	1794	SERERDRDRDREREKESILSTTTVEHAPTWRPCTEOSGSGSGSGSGSGSGSRPAS	1853
Db	11596	PEEVAEEPPAEVVEEPEAPQV-----VPPKKVPEKKA-----PAV	11636
Qy	1854	HSHAHQSPISPRQTQDALQORPSVLHNTGMKGIIITAVEPSK-----TVLRST	1901

RESULT 39
O63460 PRELIMINARY; PRT; 5027 AA.
AC O63460;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Proline-rich protein (fragment).
GN PRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI TaxID:10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Ventral prostate;
RX MEDLINE=89066721; PubMed=3198617;
RA Hemschoote K., Peeters B., Dirckx L., Claessens F., De Clercq N.,
RA Heyns W., Winderickx J., Bannwarth W., Rombauts W.;
RT "A single 12.5-kilobase androgen-regulated mRNA encoding multiple
RL proline-rich polypeptides in the ventral prostate of the rat.";
J. Biol. Chem. 263:19159-19165(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Ventral prostate;
RX MEDLINE=92250652; PubMed=1577819;
RA De Clercq N., Hemschoote K., Devos A., Peeters B., Heyns W.,
RA Rombauts W.;
RT "The 4.4-kilodalton proline-rich polypeptides of the rat ventral
RT prostat are the proteolytic products of a 637-kilodalton protein

RT displaying highly repetitive sequences and encoded in a single exon.";
RL J. Biol. Chem. 267:9884-9894 (1992).
DR EMBL; M86514; AAA41955.1; -;
FT NON TER 1
SQ SEQUENCE 5027 AA; 558470 MW; ECED8CC3EF7E260E CRC64;

Query Match 3.8%; Score 508; DB 11; Length 5027;
Best Local Similarity 18.8%; Pred. No. 1.9e-15;
Matches 608; Conservative 342; Mismatches 1165; Indels 1112; Gaps 155;
Qy 42 YQHSRDYASHLSPGSIIPQRRRPSLLSEFQPGNRSQELHLRPES--HSVLPBLGKSE 99
Db 577 YQEVSDQAEYTTSTV-----SQP-----LQELTITSEAREPHHPVPOQT 621
Qy 100 MEFIESKRPRLELLDPDLRLSP-----LLATGQAGSBDL--TKDRSLTGKLE 146
Db 622 ITVHTKHP-LVIHSEQQHNPTEVTVPQDLDELTMTPQPTAGELPQTLQDSITQIIE 680
Qy 147 PVS-----PPSPHTDPELEL-----V 163
Db 681 PPTVVVGVPPIYEEVTQTSQDAEYPPSPVSFQSLDLGLITIPTEHFIQTQTVF 740
Qy 164 PPLRSKEBLQN-----MDRVDREITWVEQOISKLKKQOQLEEEAAKPPPE 211
Db 741 PPMYTDVTLPOQSVQHLKPTGIVQPLDLELITTPQTPPE-GEISQTVQESTTQNKEPH 799
Qy 212 KP-VSPPIESKHSRLVQIYIDENRKAEEAHRILEGLGP-QVELPLYNQPSDTRQYHEN 269
Db 800 KEVAVPV-----YQAVTVTPSOYQAEVQ--KSLQPLDLELTVTSEPT-KEAYHST 849
Qy 270 IKINQMRKKLILYPRRNHARKQWKQFCORYD-----OLMEALEKKVE 314
Db 850 ISKNSLAINQVHIQHPNPA-----EATVQPLDLELTISSLSLOPTAEGELLYNQETVT 904
Qy 315 RIENPRRAKESKVEYREYKQFPIRQRELOERMQRVQSGLSWSAARSEHYSE 374
Db 905 QISPPKQV--TPPEYQEVAVP-----APVQDAKYPL---SSVLSNLSLQDELTLSS 954
Qy 375 IIDGLSEQENLEKQWQLAVIPPM-----LYDADQORIKFINNGLMADPMKVIKDRQWN 430
Db 955 ELLGEAHLQTPD---ETWVLPKDRQGIY-PDHDHKKHLNLTETVNPQPHLEHTVQHP 1010
Qy 431 MWSQEKETTFREKFMQ-----HPKNFGILIASFLERK 461
Db 1011 TIEBERSQIOKKTQITPEPKVVLQAESEEVITPILKETAPPTPHSMALQSLDEK 1070
Qy 462 -----TVAECVLYY-----YLTKNEN 478
Db 1071 LTIHSPGWTQOHANLKESKGTGKILLDYAEPNMEIELKHHGLFLKLTTEATTESN 1130
Qy 479 YKSLVRRSYRRGRKSQQOQQOQQOQ--OQOQQOQPMRPSOBEKDEKEKEAEKEEKP 536
Db 1131 TNQMTKSLQVTAFTQNKXSMALPQVQESQPPNMSLQPLDQ-ELTLSSQPHGWVP 1189
Qy 537 EVENDKDLKEDTDDTSGEDNDEKAEVASKGRKTANSQGR-----KGRITRSMAN- 588
Db 1190 HIPNTEKILYHAEAPPTGPFVEPPELFLTKTKSRPVQGTATQAAASPKEWVSRAPEK 1249
Qy 589 -----EANSERAITPQQAELASMEINSSR---WTEEMETAKGLLEHGRNWSAI 637
Db 1250 EAVLSGPGEDQESPPNMSLQSLDQELTLSSQPHGIHPNPN-----HGKILYH 1302
Qy 638 ARMVGSKTVSQCKNFYFNKQNLDELIOQHLKMEKERNARKKKKAPAAAEAEAF 697
Db 1303 AEPTGPFVEPDLFFLTKTKSPVETLIR-----TKSRKEMVSQPKYEEAVLP 1354
Qy 698 PVEDEMEASGVSGNEEMVEEALHASGNEVPRGECGPATVNNSSDTESTIPSPHTE 757
Db 1355 -----VHGEQEESSR-----SPNMSLQSLQEQ 1377
Qy 758 AAKDTGQNG--PKPATLGA-----DGPFGPPTTP-----RRTSRAPTEPT----- 797
Db 1378 LTLSSQPHGWVPHPNTHGKIYLYHAEPPPTGFVEPDLFLRTTKSPVQGTATRMVKS 1437

Qy 798 -----PA-SEATG-APTPP-----PAPP-SPS- 816
Db 1438 PEEMVSLDPENKEAVFPAQGEKGEPSPSNMSLQSLDHELFMSSQPHGWIPHPKPTDK 1497
Qy 817 -----APPVVVKEKEE-----ETAAAPVVEGEBEQKPPAAEELAVDTGKAEEPVKSECT 867
Db 1498 IYLYHAEAPPTGPFVEPDLFLRTTKSPVQGTATTEMAKSPKEMVSOTPEYKEAVLSGPG 1557
Qy 868 EAESEGA-----KGDAEAEAT-----ASGALKAEKKEGSGR----- 902
Db 1558 EDQESGPPNNTSLKSLDQEVAMSSQPHSGVPHPPKTEGKIYLYHSIEPPGPFVFKPTDLI 1617
Qy 903 --ATTAKSSGA-----PQDSNSATCSADEVDEAGGDKNRLLSRP 942
Db 1618 LVKTTTKSPAETPRRIDKLLKEMVPHSPYEAEAVFPAHGEQDES--GSPNMPQLPLD 1676
Qy 943 SLLPTGDPRA-----NASPQ-----KPLD--LKQLKQRAAAIPIQVT 979
Db 1677 QELTSSQPHGWVPHPNTPGKIYLYHAEPTGPFVEPDLFLRTTKSPVQGSPEIA 1736
Qy 980 K-----VHEPREDA-----PTKPAPPAPPP-----QNLQESDAPQPGSS 1018
Db 1737 KSPKEMVSOTPEYKEAVLSGPGEDQESPPNMSLKSLOQEVMTSSQPHSGVPHPP--KT 1795
Qy 1019 PRGK-----SRSPAP-----PADKEAFAAEAOQLPGDPPCWTSGLPFPVPPREVIK-----A 1065
Db 1796 P-GKIYLYHSIEPPGPFVKPTDLILVTKTKSPA-----WT-----PRRIDKLLKEM 1843
Qy 1066 SPHAPDSAFSAPPGHPLPLGLHDTPVLPRTPTISNPPPLISSAKHPSVLERQIGAI 1125
Db 1844 VPHSPE-----YBEAVFP-----AHGEGQDESQSPNMPQLP-----LDQELTSL 1883
Qy 1126 SOGMSVQLHVP-----YSEHAKAPVGPV-----TMCGLPLM--DPKKL 1161
Db 1884 SQPHGWVPHPNTPGKIYLYHAEPTGPFVEPDLFLRTPKSPVQGTATQMAKSPSEM 1943
Qy 1162 APFSGVQEQLSPPRGAGPESGLVPTAQEASVLRGALGSPVGGSIITGPISTRVPSDS 1221
Db 1944 VLSLSPKNETVFPAGKGQDESPPNLSLQSLDQELTMSQPHGWIH-----PPN- 1995
Qy 1222 AITVRGSI THG-----TPADVLYKGTITRIIGEDSPSLDRGREDLSKPG 1266
Db 1996 -----THGKIYLYHAEPTGPFVEPDLFLTKTKSPKMQSPRQIDKSPKEMFTQ- 2046
Qy 1267 HVIYEGKKGHVLSEGGMSVTQCSKEDGRSSGPHETAAPKRTYDMMEGRVGRASAS 1326
Db 2047 -----SPYEESLLPAHAEQOESRAPPHFSLOP-----LDQELSLSS 2084
Qy 1327 IEGLMGRAIPPERHSPHHLKEQHHRIGSITOGIPRSVVEAQEDYLRRREAKLLKREG--- 1382
Db 2085 -----HPHGWIPHPNTPDKIYLYHAEPTGPFVEPDLFLTKTKSPVQGTATK 2135
Qy 1383 TPPPPPSRDLTEAYKTQAL-GP-----LKLKPAHEGLVATVKEAGRSIH--E 1427
Db 2136 TDKSPEDRVSTPEYKEAVLSGPGEDQESPPNMSLQSLDQELAISSQPHGWIHPSPN 2195
Qy 1428 IPRELBHTPELPLAP-----RPLKEGSIITQ-----GTPLYKDTG 1462
Db 2196 APDKIYLYHAEPTGPFVEPDLFLTKTKSKPL--QGTPTQMAKSPKEMVSOTPEYKAD 2254
Qy 1463 ASTTGSKKHVRSILGSGRTPFPVPHLDVNMADARALERACYEESLSKSRP---GTASSS 1518
Db 2255 LSAGENRDE-----SPSPNMSLHPLD-----QELSLSSQPHGWIHPNPT 2296
Qy 1519 GGSII-----ARGAPVIVPEL-----GKPRQSPLTYEDHGAPFAGHLPRGSPVTWRE 1564
Db 2297 HGKIYLYHAEAPPTGPFVEPDLFLTKTKSPVQGSQIDK-----SPKEVFS 2345
Qy 1565 PTPRLQESLSSSKASQ-----DRKLTSTPREIAKSPHSTVPEHHPH-PIS 1609
Db 2346 QSPSESESVLPAQAEQOESRAPPHMSLQSLDQKLT-----LSSHPHGWIP-HHPNTPGK 2399

QY	1610	PYEHLRGVSG-----VDLY-----RSHIPLAFDFTSIPRG-----IPLDA	1645
Db	2400	IYLYAEPPPTGPFVEPPDLFLTKTKSKPVQSGPRQVDKPKEMFTQSPYEESVLPQAQ	2459
QY	1646	AA--AYYLPRHLAPN-----TYPH-----LYPP-----YLIRGVPDTPAALENRQTII	1686
Db	2460	EQEESRAPPHMSLOPLDQDLTLSSHHPGHGITHHPNTPDKIYLHYAEPPTGPFVEPPDLF	2519
QY	1687	NDYITSOQMHNTATAMAQADMLRGLSPRESSLALNYAGPRGIDLSQVPHLPVLVPP	1746
Db	2520	FLRTTKSKPAQWTPTQMAKSPPEMVLSPENKETV--FPAQGGQEESSISPPHMSLOPLD	2577
QY	1747	TPGTPATAMDLAYLPAPQPFSS--RHSSPLSPGCP-----THLTKPTTSS	1793
Db	2578	QDLTPSSHHP--GWIPHPNTHGKIYLHYAEP--PTGPFVEPPDLFLTKTKSKPVQSP	2633
QY	1794	SERDRDRDRDREREKSILTTTVEHAPIWRPGTEQSSGSSGSSGGGSSRRPAS	1853
Db	2634	RQIDKSPKEVFTQSPYEESVL-----PAQAEQEE-----	2664
QY	1854	HSIAHQHSPISPRQD--ALQORPS--VLHNTGMKGII-----TAVEPSKPTVLRS	1900
Db	2665	-SRAPPHMSLOPLDQDLTLSSHHPGHGITHHPNTHGKIYLHYAEPPTGPFVEPPDLFLTK	2723
QY	1901	TS-----TSSP-----VRPA-----ATPPP-----A	1916
Db	2724	TKSKPVHSGPRQIDKSHKEMTQSPYEESVLPQAQGEESRAPPHMSLOPLDQDLTL	2783
QY	1917	THCPLG-----GTLDGVIYPTLMEPVLPLP-----KBAPVARPERPRADTGH	1957
Db	2784	SH-PHGWPHPNTPDKIYLHYAEPPTGPFVEPPDLFLTKTKSKPVQSGPRQIDKSPKE	2842
QY	1958	APLAKPPARSGLPASPSPK--GSEPRPLVPVSGHATARTAKNLAPHA-SPD---	2010
Db	2843	MFTQSPYEESVLPQAQDOESRAPPHMSLOPLDQDLTLSSHHP--HGWIPHPNTPDKIY	2901
QY	2011	-----PPAPPASADP--HREKTQSPFSIQELELSRLGVHGS-----SVSP	2050
Db	2902	LHYAEPPTGPFVEPPDLFLTKTKSKP-----VHGSPRQIDKSPKEVFTQSP	2948
QY	2051	EGVEPVSVPSPSLTHDKLPKH--LEELD-----SHLEGEL--RPKQPGFVKLGGEAA	2101
Db	2949	EYEEESVLPQAQDOESRA--PPHMSLOPLDQDLTLSSHHPGHGITHHPNTPDKIYL---	3002
QY	2102	HLPHURPLPEQSPSSPLLOTAPGVKGQRVVTLAQHLSSEVITQDYTEHHQQLSAPLPA	2161
Db	3003	HYAEPPTGPFVEPPDLFLTKTKSKPVHSGPRQIDKSPKEVFTOS-----PEYEEVSLPA	3057
QY	2162	PLYSFPGASCP-----VLDLRPPPSDLYLPPPD	2189
Db	3058	QAEDQESRAPPHMSLOPLDQDLTLSSHHPGHGITHHPNTHGKIYLHYAEPPTGPFVEPPD	3117
QY	2190	-----HGAPARGSP-----HSEGGKRSPEPNKTSV--LGGGED-----GIEPV-	2225
Db	3118	LFLLRTTKSKPVQWPTQIDKSRSEEMVSGSPENEEADIPRHGDGQDELNRPPIISLOPLD	3177
QY	2226	-----SPEG--MTPEGHSRSVAVPLLYRD--GEQTEPSRMGSKSPGNTQPPAFFSKL	2275
Db	3178	QBLTLTSSQPHGVVPPPTPDKIY--LHYAEPPTGPFVEP-----PDLFFLRT	3223
QY	2276	TESN-----SAMVSKSQEINKLNTNHRNPEYFN--ISQPGTEIFNMFAITGTGLMT	2326
Db	3224	TKSKPVQGTPTQAKSPKEMVQ-----TPYKEAVLSAPGEDQETP-----	3266
QY	2327	YRSQAVQEHASTNMGLEAIRKALMGKYDQ--WEESPP-----LSANAFN-----	2370
Db	3267	-----SSPNMSLOSLDQBLTSSQPHGLPHPPNTHGKIYLHYADPPTGPFVEPPD	3317
QY	2371	---LNASASLPAAMPITADGRSDHDLTSPGGGKAKVSGRPSRKAKSPAPCLASGRDP	2427
Db	3318	LFLLTKTKSKPVQGTQVQMD-----KSPKQGL-----	3345
QY	2428	PSVSSVHSEGDGCDNRRTPLTNRVME--DRPSSAGSTPF-----PYNPLINRLQAGVNASPP	2480

3346 ----FAHGEAQDESSPPNMSLOPVDQELSLSSQPHGMITHPNTPGKIYHL---YABPP 3398

2481 -----PPGL-----PAGSPLAGPHAWDEEPKPLLCQYETLSDSE 2517

3399 TGFVEPPDLFLTKTKSKPLAGSPQMAKSPKEMPS---QTLHKE 3442

RESULT 40

O88737 PRELIMINARY; PRT; 3942 AA.

AC O88737;

DT 01-NOV-1998 (TREMELrel. 08, Created)

DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Bassoon.

GS BSN OR BASSOON.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129 SVJ;

RX MEDLINE=98345363; PubMed=9679147;

RA Dieck S., Sanmarti-Vila L., Langnaese K., Richter K., Kindler S.,

RA Soyke A., Wex H., Smalla K.H., Kampf U., Franzer J.T., Stumm M.,

RA Garner C.C., Gundelfinger E.D.;

RT "Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively

RL J. Cell Biol. 142:499-509(1998).

DR EMBL; Y17034; CAA76598.1;

DR EMBL; Y17035; CAA76598.1; JOINED.

DR EMBL; Y17036; CAA76598.1; JOINED.

DR EMBL; Y17037; CAA76598.1; JOINED.

DR EMBL; Y17038; CAA76598.1; JOINED.

DR PIR; T42730; T42730.

DR MGD; MGI:1277955; Bsn.

DR GO; GO:0045202; C:synaptic junction; IEA.

DR GO; GO:0046872; F:metal ion binding; IEA.

DR InterPro; IPR008899; Znf_piccolo.

DR Pfam; PF05715; Zf_piccolo; 2.

SQ SEQUENCE 3942 AA; 418739 MW; 150267E636C4DACB CRC64;

Query Match 3.8%; Score 504; DB 11; Length 3942;

Best Local Similarity 19.1%; Pred. No. 2.1e-15;

Matches 590, Conservative 324; Mismatches 1122; Indels 1052; Gaps 136;

QY 55 PGSIQQRRRRLSELSEFPQGNERSQELHURPESHVYLPGLKSEMEFIESKPRLELLP 114

Db 557 PQGLGQPSGLPAKAS-----POATKASPOATKASPOATKASPOATKASPP 601

QY 115 DPLLRPSPLATGOPAGSEDLTKDRSLT-GKLRPVSPSPPHPTDPELELVPRLSKEELI 173

Db 602 ----QAKPLRATPEKTSKSSAOKKATPAKAPV-PKPPET-----TVPCTPKAK-- 649

QY 174 QNMDVRDREITVBEQKISLKKKQQQLLEBAAPKPEPEKPVSPPIESKHSLSVLIYDE 233

Db 650 SGVKRTDPAIPVVKP-----VPEAPKGEAEPEVPKPYSDLSRS----- 689

QY 234 NRKKAEAHRILEGIGLGOVELPLYNQSPDTFQHENIKINQAMRKLLILYFKRNHARKQ 293

Db 690 -----PQ-----SLSDTGYSDGVSQS----- 708

QY 294 WKQFCQRYDQLEALEKVKVERIEN--NPRRRAKESKREYVEYKQFPEIRKQRELOERM 350

Db 709 -----EITGVQVEQVQLDSAGVTGPRPPSPSELHKVGSLSRPSLEAQAVPSAE 758

QY 351 QSRVQCGSGSLMSAAKSEHEVSEIIDLSEQENLEKQMRQLAVIPMLVDADQORIKFI 410

Db 759 WSKPPRSSSSSAVEDQKRRPHSL-----IMPEAFDSDELDGIL 797

QY 411 NMGLMADPMKVKYKDRQVMNMNWSQEKETPREKFMQHPKPNGLIASFLERKTVAECLVY 470


```
Qy 2088 PKQPGVKLGEEAAHLP--HLRPLPESQPSSPLLQTPAGVKGHQVRVTLAQH----- 2138
Db 2769 KKKDPLEIGVQ-AHLPEESLSQLVSRQPKSPQVLYSP-----VSLSPHRLDTSF 2820
Qy 2139 -ISEVITQDYTRHHQOLS-----ADLPAPLYFPFGACPCVLDLRPPPSDLYLPPP--- 2188
Db 2821 ASSERLNKAHVSPQKFIADSTLRQOTLPRPMKT-----LQSLSD---PKPLSP 2867
Qy 2189 --DHGAPARGSPHS-EGGKRGPEPNKTSVLGGGSDGIE-----PVSPPGEMTEP--GH 2236
Db 2868 TAESAERFSLYHQGGLS-----QVSVL--PPNGLVRKVRTLPSPFPPEEAHLPLAQ 2921
Qy 2237 SRSVAV--PLLYRDGEQTEPSRMGSKSPGNTSQPPAFPSKLTESNMAVKSQKQEIINKL 2294
Db 2922 VPSQLYAAASLLQR-----GLAGPTTVPATKASILLRELDRLRLVEHESKLRKQ 2971
Qy 2295 NTHNRNPEVYNISQFTEI-----FNMPATGTGLMYRQAOVQEHAS-- 2337
Db 2972 AELDEEKEIDAKLKYLELGITORKESLAKDRGDRYPLRLGLG-----EHRDYL 3021
Qy 2338 -----TNMGLEALIRKALMGKYDOWEESPPLSANAFNP--LNASASLPAAPIITAADGRS 2390
Db 3022 SDSELNQLRGCTTPA--GOYVDYPASAAVPATPSGTPTAFQQRFPFPAPQYTAGS--- 3076
Qy 2391 DHTLTSFGGGG---KAKVSGRPSRRAKSPAGLASGDRP--PSVSVSVHSEGDGDRTP 2444
Db 3077 ---SGPTQNGFPAHQAPTYYTGPSTYPTATYPPGTPGYPAPGLSPQAFHTGHYAATP 3132
Qy 2445 L-----TNRVWEDRPSAG-----STPPFYNPLMRLQAGVWASPPPPGLPAGS 2488
Db 3133 MPTTQSAFPFVQADSRRAHQPRQTSADLQKQVNTYEVIGSPAVTMSSAPPE--TCYS 3190
Qy 2489 GPLAGPHAWDEEPKLLCSQVETLSDS 2516
Db 3191 GPAVSGSYEQGAKEPHRGDRSSVSQS 3218

RESULT 41
Q10465 PRELIMINARY; PRT; 7962 AA.
ID Q10465;
AC Q10465;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Elastic titin (Fragment).
GN TIIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96026330; PubMed=7569978;
RA Labelit S.; Kolmerer B.;
RT "Titins, giant proteins in charge of muscle ultrastructure and
RT elasticity."
RL Science 270:293-296(1995).
DR EMBL; X90569; CAA62189.1; -.
DR PIR; I38346; I38346.
DR GO; GO:0030017; C:sarcomere; TAS.
DR GO; GO:0008307; F:structural constituent of muscle; TAS.
DR GO; GO:0006941; P:striated muscle contraction; TAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR004168; PPAK_motif.
DR Pfam; PF00047; Ig; 59.
DR Pfam; PF02818; PPAK; 53.
DR SMART; SM00408; ICG2; 43.
DR PROSITE; PS50835; IG_LIKE; 58.
FT NON_TER 1
FT NON_TER 7962
```

```
SQ SEQUENCE 7962 AA; 883018 MW; B85240533CBADES8 CRC64;
Query Match 3.8%; Score 503.5; DB 4; Length 7962;
Best Local Similarity 19.0%; Pred. No. 5.5e-15;
Matches 534; Conservative 355; Mismatches 1053; Indels 865; Gaps 128;

Qy 27 PVQIARTHTDVGLEVQHHSRDYASHLSFGSIITQPORRRPSLSEFPQGNERS-----Q 80
Db 5498 PIQFTRIQNIIVSE--HOSATFECEVSDDAIVTWYKGTELTESQKYNFRNDRCHYM 5555
Qy 81 ELA-LRPESHSLPELGKSEME-----FTEKRPRLLEL---LPD----- 115
Db 5556 TIHNVTPDDGVSIVTARLEPRGEARSTAEIYLTKEIKLEKPPDIPDSRVPIPTWPIR 5615
Qy 116 -----PILRPS-----LLATGQFAGSED-----LTKDRSUTGKLEPV 148
Db 5616 AVPPEEIPPVAVPVPVLLPTPEEKPPKRIEVTKKAVKDAKKVAVAKEMTPREIV 5675
Qy 149 SPSPDPTDPELELVPPRLSKBELIQNMDRVREITWVEQOISKL-KKQOQOLEEBAKP 207
Db 5676 KQPPPTT-----LIPAK--APEIIDVSSKAE-----EVKIMTITRKKEVQKEKAV-- 5720
Qy 208 PEPEKPSVPPISKHSRLVQIIVDENRKAAEAHRLLEGLPQ-----VELPLYNQPSD 262
Db 5721 -----YEKK-----QAVHKERVFIEFEPYDELEVEPYTEPEQPYVEBDE 5764
Qy 263 TROYHENIKINQAMRKLI-----LYFKRR--NHARKQWKQKFCOR----- 301
Db 5765 D---YEEIKVE--AKKEVHEEWEEDFEEQYVVEREEGYDEGEWEAEVQEREVIQVQ 5818
Qy 302 ---YDQUME-----ALEKK-----VERIENPPRRAKE-----SKVREYIEKQ 336
Db 5819 KEVYESHERKVPKAPVEKKAPPPKVPKIPKVPKIEKTSRRMEEEKVQVTKVPEVSKKI 5878
Qy 337 FPEIRKQRELOERM-----QSRVGORGSLG---MSAARSEHEVS---EIIDGLSQEN 384
Db 5879 VPQPSRTPVQEEVIEVKVPAVHTKQVISEEKQFPFASHTEESESVTVPEVQKEIVTEEK 5938
Qy 385 LEQMRLQAVIPMLYDADQ-----QRIKFNNMGLMADPMKVY-----KDRQVMMWSEQE 436
Db 5939 IHVAVSKRVEPPPKVPPELPEKPAPEEAVPPI-----PKKVEPPAPKVPPEKVPPEE 5992
Qy 437 KETP---REKFMQHPKNGFLIASFLERKTVAECVLYLYLTQKQENYKSLVRSYRRRCKS 493
Db 5993 KKPVPVPKKEPAAPPK-----VPEVPKVPPEEKIPVPVAKKKEAPPKAPVPEV--QGVV 6045
Qy 494 QQQQ-----QQQQ-----QQQQQQQQQMPRSSQE-----EKDEKEKEKE 528
Db 6046 TEKQITIVTQRESEPPPAVPEIPKKVPPEKVPKPEEVEVPPPPKVPALPKPVPPEKV 6105
Qy 529 A-----EKBEKEPEVNDKEDLLKEKTDDTSGEDNDEKAVASKGRKTANSQGRKGRIT 583
Db 6106 AVPVVAKKAPPAPPAEVSKTTVV-----EKRFAVEEKLSPFVQ-----RVEVT 6150
Qy 584 RSMANEANSEEAITPOQSAELASLWEL-NESRRTTEEMETAKGLLEHGRNW-----SA 636
Db 6151 R---HEVSABEEWSYSEEBEGSVISVYRBEEREBEAEVTEVMEPEPEYVVEEKLHI 6207
Qy 637 IARMVSGK---TVSCKNFYFNYKKQNLDE-----ILOQHLKM---EKEK 677
Db 6208 ISKRVAEAPAEVTERQEKIIVLPKIPAKIEEPPPAKVPPEAKPILVPEKKVPAPVPKKEK 6267
Qy 678 NARRKKKKKAPAAASEEAAFP--VVEDEMEASGVSGNEEMVEAEALHAGNEVPRGEC 736
Db 6268 VPPKVPPEEKVPPEKKVPKVKMEPLPAKVTEKHMQITQEKVLVA-----VTKEA 6323
Qy 737 SGPATVNNSSDTEISIPSPHTAAKDTQNGKPKPATLG-----ADGPPGPPT 784
Db 6324 PPKARV---PEEPKRAVPEEKVLKLPKREBEPPKAVTEFRKRVVKEKVSIEAPKREPQ 6380
Qy 785 PPRTS-----RA-----PIEPT----- 797
Db 6381 PIKEVTIMEKERAYTLEBEASVQREEEVEEYDYKEFEYEPTTEEDQYBEYERE 6440
```

Qy 798 -----PASEATGAPT-PPAPSPSPAPPVVPKKEEETAAA-----P 835
Db 6441 YERVEEHEEYITEPEKPIPVKVPVEEPVTPKPAKVLKAVPEEKVPVPIPKKLKP 6500
Qy 836 PVEEGEQKPPAAEELAVDTKAEPEKSECTEAEAGPAKGADEAAEATGALKAOK 895
Db 6501 PPKVPEPKKVFEEKIHISITKRE---KEQVTEPAKVPKPK-----RVVAEKKVPVR 6552
Qy 896 KEGSGGATTAAGSGAPQDSSATCSADE-----VDEAGGDKNRLSPRSLIT----- 946
Db 6553 KE-----VAPPVVPVEPKLEPEEVAPEEVEVTVHVEEVLVEEYIEHEEPTTEEVV 6608
Qy 947 ---PTGDPRAWASP-----QKPLDKQLKQRAAAIPPIQVTKVHPPEPREDAAPTKAPAP 999
Db 6609 PVIKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6664
Qy 1000 PPNQLOPESDAPQOQSSPGKRSRSPADKAEFAAEAKOQLPGDPPCWTSGLPVP 1058
Db 6665 PP-----AKVPEVP-KKVPPEKVPVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6706
Qy 1059 PREVIKASHPADPSAFSAAPPCH--PLPLGLHDTARPVLPRPTISNPPPLISSAKHPS 1116
Db 6707 EKKVPVAPKVP-----APPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 6759
Qy 1117 VLROIGAIGMSVOLHVPYSEHA--KAPVGPVTMGLP-----LPMDPKKLAPFGV-- 1167
Db 6760 -----VPPVALQPEEVLPEEVEIPEEVLPEEVEVLPEEVEVLPEEVEVL 6807
Qy 1168 KQQLSPRQGPESIGVTAQASVLRTALSGVPGGSITKGI--PSTRVPS-----DS 1221
Db 6808 EEEIPEEVEEVEEVEEVEEVEE-----FVPEEVLPEEVPKVPKVPKVPKVPKVP 6857
Qy 1222 AITVRSI---THGTADVLKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHV 1277
Db 6858 KVTCKKVPKKEAPKVP-----PEVKKVEER-----IILPKGEEV 6897
Qy 1278 LSVEGMSVTQCSKEDGSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAPP 1337
Db 6898 LPVE---VTEEPPEEPISEIEEPPSIEVEEVAAPPVPEVIKKAPEA--PTVPK 6951
Qy 1338 ERHSPHLKQHQHRTGSIQIGIPSYVEAQEDYLREAKLLKRGTPPPPPPSDLTEAY 1397
Db 6952 KVEAP-----PAKVSXKIPEKVPVPVQ--KGAAPKVPKVPKVPKVPKVPKVP 6998
Qy 1398 KTOALGLKLPKPAHEGLVA-----TVKE-----AGRSIHETPRELHRTPELPLAPR 1444
Db 6999 KGAAPKAGRTVLEEKVSVAFROEVVVKERLEEVVEAEVEEIEPEEPEHEV-----E 7053
Qy 1445 PLKEGSITQGTPLKYDTGASTGSKGHDVRSLSIGSPGRTTTPPVHPLD-----VMADARALE 1500
Db 7054 YFEEGEP-----HEVEEFIKLEQHRVEEHRVEKVRHVIIEVEFAEE 7094
Qy 1501 RACVEESLKRPGTASSGSGIARGAP--VTVPLGKPRSP-----LYVEDHAGAPAG 1552
Db 7095 VEVEKEP-KAPPKGEISEKIIIPKKGPTKVPKPEPAKVPKVPKVPKVPKVPKVP 7150
Qy 1553 HLPKGSVTVREPTPLQEGSLSSKASQDKLTSTPREAKSPHSTVPEHHHPHPIGPE 1612
Db 7151 BEPVPTKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 7209
Qy 1613 HLLRGSVDLYRSHIPLAFDPTSIPIRGIPLDAAAAYVLPRLHAPNPTPHLYPPYLIRG 1672
Db 7210 -----KPEVP-----PTKVPB-VPKAAVPEKKVPEAIIPKPEP-----PPEVPE- 7248
Qy 1673 YPDTAALENQTIINDYITSOQMHNTATAMAQADMLRGLSPRESSLA-----LN 1723
Db 7249 -----LPVLVP-----EPEEPPSAPPKKPEVPPVR 7267
Qy 1724 YAAGPRGIIDLQVPH-----LPVLVP-----PTGCTATAMDRLAYL 1761
Db 7268 VPEVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 7327

Qy 1762 PTAPQPFSSRHSSSPLSGGTHLTKPTTTSSRERDRDRDRDREREKSIIITTTV 1821
Db 7328 PKVAVPEKKVPEAIIPKPEP-----PPEVPEPEVEALEEPAEVEPEPAAPQVTPV 7383
Qy 1822 EHAIWIPGTTEOSSGSSGSSGSSRPAASHAHQHSPISPRTQDALQORPSVLHNT 1881
Db 7384 PKNPV-----PEKKAPAVAKKPELPEV-----KVPEVPKEV 7415
Qy 1882 GMKGIIITAVPSKP-----TVLRSTSTSSVVRPAATPPATHCPLGGTLGQVY 1929
Db 7416 VPEKKVPLVVPKKEAPEAKVPEVPKVPKVPKVPKVPKVPKVPKVPKVPKVP 7463
Qy 1930 PTLMEPVLLPKAEPRVARPERPRADTGHAFIAKPPARSGLEPASPSK-----GSEPRP 1983
Db 7464 EVPKKPVL--EEKDAVPVPERAESPPPEV--BEPEIAPBEETIAPBEKVPVVAESEP 7519
Qy 1984 LVPPVSGHATITARTAKNL-----APHASDP-----PAPPASADP 2021
Db 7520 EVPP-----PAVPEBPKIIPKVPVVIKKPEAPPKPEPEKVPKVPKVPKVPKVP 7575
Qy 2022 HREKTQSKPFSI-----QELRLRSLGVHGS-----SYSPGVEPVS 2057
Db 7576 PKEDVKEKIQLKXAIKKKVPENQVPEKVELTLPKVPGEKKVKVKLLPERKPEKEVV 7635
Qy 2058 PVS-----SPSLTHDKGLPKHLELDKSHLEGELRPQKQPVK-----LGGEAAHLPHLR 2107
Db 7636 LKSVLRKRPEEPEKVPKLEKVKKPAVP--EPPPKPEVEVEVPTVKRERKIPEPT 7692
Qy 2108 PLPESQSSSPLLOTAGVKGHORVTVLAQHISEVITQDTRHHPOQLSAPLPALYSFP 2167
Db 7693 KVPEIKP-AIPLPAPEPKPEAEVKT-----IKPPVEPEPTPIAAEPTVPVVGKK 7743
Qy 2168 G-ASCPVLDLRRPDLVPLPDHGAARGSPHSEGGKRSPEPNKTS--VLGGGEDGTEP 2224
Db 7744 AEAKAPEEAAKPK-----GFIKVP-----KKTSPIDEAERKLRPGSGGKPK 7787
Qy 2225 -----VSPPEGMT----- 2232
Db 7788 PDEAPFTYQLKAVPLKVFKEIKDILTESBFGSSAIFECVLSPTAITTWKDGSNIRE 7847
Qy 2233 EPGHSRA-----VYPLLRYDGEQTEPSRSGSKSPGNTS-----QPPAFSKLITES 2278
Db 7848 SPKHFRTADGKDKRLHIDVQLSDAGEYTCVLRGLNKEKTSKLVVEELVRFVKTLEE 7907
Qy 2279 NSAMVSKS---KQEIKNLKNTHNRNPEVYNISQGTGFIPNMPAITG 2321
Db 7908 ETVVVKQQLVLSCLNKRDRDVRKDGKIVVEKGRIV---PGVIG 7951
RESULT 42
Q9N3Y8
ID Q9N3Y8 PRELIMINARY; PRT; 2344 AA.
AC Q9N3Y8, 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Y40C5A.3 protein.
GN Y40C5A.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

RA Kalicki J., Randall J.;
RT "The sequence of C. elegans cosmid Y40C5A.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024772; AAF60538.1;
DR WormPep; Y40C5A.3; CE24257.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
SQ SEQUENCE 2344 AA; 245902 MW; 7D8696D48D0ADDBB CRC64;

Query Match 3.8%; Score 503; DB 5; Length 2344;
Best Local Similarity 19.2%; Pred. No. 1.2e-15;
Matches 461; Conservative 275; Mismatches 865; Indels 796; Gaps 109;

Qy 515 SSOEKEDEKEKEBAEKEEKEPVENDKEDLLKKTDDTSGDENDEKEAVASKRKTANS 574
Db 218 STTEFTTEETTVSTNPPTPELSDD-----PTDNDATDIDDK----- 257

Qy 575 QRRKGRITRSMANESSEAITPQQAELASLMELNESSRWTEEMETAKGLLEHGRNW 634
Db 258 -----IVKVESSTYTAETVAPFFENTVTSVFEEPPDDADFDKPVKSKSTLPNGINS 310

Qy 635 SAIRMVGSKTQCKNFYFYKRRQNLDEILOQHLKMEKERNARKKKKAPAAASEEA 694
Db 311 KRSTSYANPLTVDNA-----AABSRNKLSCPLCAPGTV 344

Qy 695 APPVVEDEMEASGVSNDEEMVEEAEALH--ASGNEVR-----GECGPATVN-NS 745
Db 345 AVPP-----KEENLPPIEPLPSVAGYQPPRTLPVAPCDPSESTFVQIHE 389

Qy 746 SDTISIPSPHTEAAKQTQNGKPKPATLADGPPPTPPRTSRAPIEPTPASEATGA 805
Db 390 TETSPASPSPSSVSVTIPRVLPASVSPSPASVNPVTIILSAPAK-----SPATYG 445

Qy 806 FTP-----PPAPPSAPP-----PVVPKE-EKEEETAAPPVEEKEKPPAA----- 848
Db 446 PAPEPSRPAPVPVAPAPSVIPTTSQSPQKQTPAQPTSPAPKIPPEGPAQSPNAVKTSY 505

Qy 849 -----RELAVDTGAE--EPVKSECTEABE-----GPAKGD 879

Db 506 GSSPAAPRVPEMIEVAPGNVEKTPDOTNQVPETSHENAPAPQSPSSPVYGTPTSEP 565

Qy 880 AEAATATAGALKAR-----KKEGSGRATTAKSS-----GAPQDSD-- 916

Db 566 AKPSMPSGESPAQPEPSIPAVNPSPSQPSSSGPIQTAPSPSPNAIPEGPPGPDVP 625

Qy 917 -----SSATCSADEVDEAEGDKNRLLSRPSL-----LTP 947

Db 626 VMIALPPPKSPKNSGAGEDKGEVDQEVNPSER-----RPEISSISENSIENISKTTP 681

Qy 948 TGDP-----ANASQKPLDLQKORA-----AAIPPIQVTKHE----- 983

Db 682 TTVVPKSTITGRMPSVNPNSDNLSNMEDGNHINQSEFSNLSSETSITPASHFHGDGKKQ 741

Qy 984 -----PPREDAA-----PTKPAPPAPPPQNLQSPESDAPOQPGSGSPGKSRSP 1026

Db 742 EKLPGSKGPIPPROESITGSTRHTRGSSPSIIPYATAPTSPPDKKPTIP-----PTFP 798

Qy 1027 APPAKAEFAAQAQ-----KLQDPPCWTSGLPFPVPPREVIKAS-----PHADDP 1072

Db 799 TLSVDSGDIGQELRITRSPPNVALPSRPSKTLTTPSELPIITNLPSARQSTSPASVENV 858

Qy 1073 SAFSVAPPGHPLPLGLHDARTVLPVLPPTISNPPPLISSAKHPSVLERQIGAISQCMVQ 1132

Db 859 SNNSPSPGQSQIPI-----ESTVIFGTPTSPISPMETRSTAN-----SSK 899

Qy 1133 LHPVYSEHAKAPVGPVTMGLPLFMDPKLAPPSGVKQ-----EQLSPRGQAGPPESL----- 1184

Db 900 NHPVSSKISNRLPVSTGSKIHLNKT--SPOTGAMENSGEVISEEPIGKPVSPGGTG 957
Qy 1185 -----GVPTAQEASVLRGTALGVPGGSITKGIPISTR-----VPSDAITYRGSIIT-- 1230
Db 958 AALPYTSPFKTOKSSTFASTDVHSASAPSLPEKLSHKNFEKATPPSISAAITTNVVVD 1017

Qy 1231 ---HGTPA-----DVLKGTITRIICEDSPSRLDRGREDSLPKGHVYIEGKXGHVLS 1279
Db 1018 AFENSSPAAPAVSTWEEVTPSGT-TASISQSQPT-----EGPPGAVPP 1059

Qy 1280 YEGGMSVTQCSKEDGRSSGGP-PHETAAPKRYDMMEGRVGRALISASIEGLMGRALPPE 1338
Db 1060 YALFPRTHQSSNSLSTGSSLPTSSSSAPVK-----VSSKTESM--KTPPS 1105

Qy 1339 ---RHSPHLKEQHHRIGSIITQIGIPRSVVEAOEDY---LRREAKLLKREGTTPPPPSRD 1392
Db 1106 SVSSSGSLQIQEPTQAS-----SNVVEISESSSPVSSVSTWEEITTSQPSAPTQ 1159

Qy 1393 LTEAYKTOALGPLKLP-----AHEGLVATVKEAGRSIH---EIPREELRHTPELPLA 1442
Db 1160 ILTAKSRPTPGQGI VPPVYALPFRPHETL-STASPPASSLSASAKLPKSESQFMNAIP-- 1216

Qy 1443 PRPL-----KEGSITQGTPLKYDTGAST---TGSKKHQDVRSGLISGPGTFPPVHPLDVM 1493
Db 1217 PSEFVDSGSGSGFGSLQIPTTDSSTPMDPSMEDYDVLTVPTAVTSSPVLB----- 1272

Qy 1494 ADARALERACYBESLKSRCPTASSGGSIARGAPVIVPE-----LGKPROS 1539
Db 1273 -----TSENRSQSLITS-----LISSGKPAIPDMIMAMWTSQSSSKFSKSTAF 1317

Qy 1540 P-LTYEDHGA--FAGHLPRGSPVTRREPTPRLOEGSIS-SSKASQDRKLTSTPREIAK 1594
Db 1318 PTLITQTEIKPTTVSSFPAPKVPETKINQPTISSNIGDLSKTEKESTIPREVEVP--SA 1374

Qy 1595 SPHSITPEHH-----PHPSIPVEHLRQVGVLDLYRSHIPLAFDPTSIP 1638
Db 1375 SPBSTVTEHIDAETNSATIPTGTTPPSPLKPKTKL--GLT-----SH-PSAIPPAIS 1425

Qy 1639 RG--IPLDAAAAYLPRHLAPNPT-----YPHLYPP----- 1667
Db 1426 SKTLAPPVAPPTVTSNIAFSTTQSQOQTRPTTTHRPGITPPLAPKTIYPSSLQTS 1485

Qy 1668 -----YLIRG-----YPTDALEN-----RQTI----- 1685
Db 1486 SSPTPGTSSIIIVAGSRASSNPTTASITDGSSEEBEENRILPEETLSPASARLWT 1545

Qy 1686 ---INDYITQQMHNTATAMAQADMRLGLSPRESSLA---LNYAAGPRGI----- 1731
Db 1546 PDMISKYISSKAISTESVSS-AEQPSYLSNLPQSSSVASVKPPSSSVQNIETVTPS 1604

Qy 1732 -----IDLSQVPHLPVLVPTTGTATAMDRLAYLPTAPQPFSSRHSSPSLSPGQPTH 1784
Db 1605 YFPVQSKSVTQAPIPPASSPSAPSIPISTLPAVPS-----SSSVSPSSPAPSE 1655

Qy 1785 LTKPTTSSSERDRDRDREREKSIILTSTTVEHAPIWPGTGEQSGSGSGSGG 1844
Db 1656 -----STSTVSETP----- 1664

Qy 1845 GSSSRPASHASHAHOHSPISPT--QDALQORPSVLHNTGMKGIIITAVEPSKPTVLRLSTS 1902
Db 1665 APVNNPASNEHEAEGPYEPLSPQPAENHVEEYQNT---VDTIIEKS-----ATP 1714

Qy 1903 TSSVPRPAATPP---PATH---CPLGGLDGVYPTLMEPVLLPKEARVAPRAPERPRADTG 1956
Db 1715 APAPVTQAVQAPGPGVEHRYEIPAPGPAPG--PAL-EPAPAPTSAPQIVEP----- 1763

Qy 1957 HAFIAPKPARSGLEPASPSKSGSEPRPLVPVPSGHATIAIPAKNLAPH-HASPDPPAPP 2015
Db 1764 ---LPPVQD--LPQOPTPEPEPLPIA-----TPAQOPTHTYGAOGPNIVP 1805

Qy 2016 ASADPHREKTSQKPFSTQLELRLSLGVHSGSYSGEVPVSPSSPSLTHDKGLPK--- 2072


```

Db 1806 VTTAPVVPQETQAPI-----APSNPEVPQP--NPST--VSIPEPTG 1844
Qy 2073 HLEELDKSHLGEELRKPQGPVKLG-----GEAHLPLRLPESQPSSSPLLOTAP 2124
Db 1845 NIEEAE--HVDA--KPSYFGSSNNLEEDHEEGHIAELPALEPV--SVPTAARSL--P 1895
Qy 2125 GVKGHQVVTLAHISEVITODY--TRHHPOQLSAPLAPLYPFPQASCPVLDLRRPPSLD 2183
Db 1896 GESNYSN-----LEEDHDDKHLVGGPSVPSTPTQTSNGSSYSILEEDNQETN- 1943
Qy 2184 YLPPDHGAPARGSPHSGGKRSPEPNKTSVLGGEDG--LEPVSP-PEGWTEPGHRS 2240
Db 1944 -----NSGGPSYSPDKQVQVVPVTKPA--GGYRDGGMVENTAPQTTVQRP----- 1990
Qy 2241 VYLLYRDGEOTSPRMSKSPGNTSQPPAFSKLTESAMVSKSQKEINKLKLTHNRN 2300
Db 1991 -HPIKVVRPATKPYR-PRSPAPRIQRPY-----NPQPVVVRPQRPQRPQRP 2041
Qy 2301 EPEYNISQTEIFNMPAITGLMTYRSQAVQEHASNTMGLAIIKALMGKYDQWES 2360
Db 2042 QPQ---PQPPQ---PQPPQ---PQPPQ---PQPPQ---PQPPQ---PQPPQ---PQPPQ--- 2073
Qy 2361 PPLSAN-----AFNPLNASASLPA-----AMPITAADGRSDHTLTSPGG 2400
Db 2074 PPIFPQRPAPFIPFPQVQRPSPAGCGQLFSQGGGLCPISFNPCCQQQQQNNCGG 2133
Qy 2401 GKAKVSRGPRSKAKSPAGLASGDRPPSVSVHSEGCNRTPLTNVWEDRPSSAGST 2460
Db 2134 CGGCGSGGNSCGGCGNSGCGGCGGCGSSC-----CQPTSTATSHAANKLNRRVAV 2188
Qy 2461 PFPYNPLIMLQAGVMASSPPFGLPAGSGPLAGPHAWDERPKPLCSQYETLSDSE 2517
Db 2189 PL-----NNPFPAGINDH-----LATQWESVSEGE 2212

RESULT 43
Q81Q71 PRELIMINARY; PRT: 2465 AA.
AC Q81Q71;
DT 01-MAR-2003 (TreeRel. 23, Created)
DT 01-MAR-2003 (TreeRel. 23, Last sequence update)
DT 01-OCT-2003 (TreeRel. 25, Last annotation update)
DE CG323394-PA.
GN CG323394 OR CG10115.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC - Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN NCBI_TaxID=7227;
RP [1]
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam M.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

```

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Parasag V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradycky P., Huang Y., Kaminker J.S., Prochink S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003561; AANI2081.1; -.
DR FlyBase; FBGN0052394; CG32394.
DR GO; GO:0003824; P: catalytic activity; IEA.
DR GO; GO:0008152; P: metabolism; IEA.
DR InterPro; IPR000873; AMP-bind.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR PROSITE; PS00455; AMP BINDING; 1.
DR PROSITE; PS50014; BROMODOMAIN 2; 1.
SQ SEQUENCE 2465 AA; 267809 MW; 716D22341B181188 CRC64;
Query Match 3.8%; Score 502; DB 5; Length 2465;
Best Local Similarity 18.8%; Pred. No. 1.5e-15;
Matches 530; Conservative 365; Mismatches 998; Indels 932; Gaps 130;
Qy 129 PAGESDITKORSLTGKLEPVSPPPPHPTDPELVPRLSKKEELIQN-----MDR 178
Db 28 PGSAQQQQQQS-----QEPSSSSSSSGTGTTTTASTVRSKKQAVLNSGSSSTSKMADK 83

```


Db 1851 QQQPTQNTQSLNKKQQMFNSFLGTWAFGPIGNIAADKAFEMYNRAAMGFKQDFAKD 1910
Qy 2016 ASASDPHREKTSQSPFSSQIELESLGSLVHGSYSYSPGVEPVSPVSSPSLTH-----DKGL 2070
Db 1911 NSCOLQQQQQQQSP-----QVITNKQANQQTQQPQQSQSQSPHILQIQTALSONY 1964
Qy 2071 PKHLEELDKSHLEGELRPQKQPVKLGGEAAHLPH-----LRPLPESQPS----- 2115
Db 1965 QQQQQQTQAQKLPQQQQQQPPQQQLNYQQQAQLNHNYYTAQQCATAVDPKPPAAQSTVAAA 2024
Qy 2116 ---SSPLLQTPAGVKHQVVTLAQHISEVITQDTRHHPOOLSNAPLPAP--LVSFFCAS 2170
Db 2025 VSESSNMNLPST-AHQ-----HHLSTQ-----HHLAAYNKPTPPPPQTYSNP-LM 2069
Qy 2171 CPVLDLRRPPDLPLPPDHGAPARGSPHSBGKRSPEPNKTSVLGGEDGIEPVSPPEG 2230
Db 2070 QSMGLYAGNYFDKTMPPAAHMYSSSSAASAYGNPAQQLPGNYVPGNNPAAHQQQQQDQ 2129
Qy 2231 MTEPGHRSAYVYLLYRDGEQTEPSRMGSKSPGNT--SQPPAFPSKLTESN----- 2279
Db 2130 QQQQQQAQAPVPP-----AEVKAFAKGRKKKAATIAEAAAAAKQQQQQQQQQQQAQ 2184
Qy 2280 -----SAMVSKKQEIKNKLTNRNE-----PEYNISQPGTEIFNMPAITGTG 2323
Db 2185 QQQQVAAQQQHQQQQVSAQQQQQQHQQVAAQQQQQQHQAQMPQYNMPQ---SMAASAAAATNQ 2241
Qy 2324 LMTYRSQAVOSHASTNMGLEAIRKALMGKYDQWEEPSPLSANAFNPLNANASASLPAAWPI 2383
Db 2242 LQAH-AQAQQGFQLYAGLKS-----GGVSPVSSAATPVNSGGSTNQ--- 2285
Qy 2384 TAADGRSDHTLTSPPGGGKAKVSGRPSRKAKSPAPGLAS--GDRPPSVSVSHSEGDGCR 2441
Db 2286 TAADAAISLSTSTGG-----MVPGSAFNFAFTPTGTLGLYGOAAAASVLDQ----- 2333
Qy 2442 RTPLTNRWEDRPSAGSTPPYNNPLNRLQAGVNASPPPPGLPAGSGPLAGPH-HAWDE 2500
Db 2334 -----FRDAP-----NPYYM-----PPAHSAGTAANPSGNAADK 2363
Qy 2501 EPKPL 2505
Db 2364 QQNPL 2368
RESULT 44
O88778 PRELIMINARY; PRT; 3938 AA.
AC O88778;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Bassoon.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley; TISSUE=Brain;
RX MEDLINE=98345363; PubMed=9679147;
RA Dieck S., Sammarti-Villa L., Langnaese K., Richter K., Kindler S.,
RA Soyke A., Wex H., Smalla K.H., Kampf U., Franzer J.T., Stumm M.,
RA Garner C.C., Gundelfinger E.D.;
RT "Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively
RT localized at the active zone of presynaptic nerve terminals.";
RL J. Cell Biol. 142:499-509(1998).
DR ENBL; Y16563; CAA76287.1; -
DR PIR; T42761; T42761
DR GO; GO:0045202; C:synaptic junction; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR InterPro; IPR008899; Znf_piccolo.
DR Pfam; PF05715; Zf_piccolo; 2.
SQ SEQUENCE 3938 AA; 418417 MW; 5BF3C230E2C71AE2 CRC64;

Query Match 3.8%; Score 498; DB 11; Length 3938;
Best Local Similarity 19.5%; Pred. No. 4.2e-15;
Matches 506; Conservative 257; Mismatches 918; Indels 918; Gaps 114;
Qy 491 GKSQQQQQQQQQQQQQQQQQQQQPPRSSQBEKDEKEKEAEKEEKEPEVENDKEDLLKKEKT 550
Db 71 GSVRRLLDPKEPLSGORATSTPTKQASATAPGRESPRET-----RA 111
Qy 551 DDTGEDNDEKAEVASKGRKTANSGGRKGRITRSMANEANSEEAITP-----QQSAEL 604
Db 112 QGLSGQ-----EAGSPRTQLQVDSRTQSGR-----SPSVSPDRGSTTSPYSVQIAPL 161
Qy 605 ASMELNBSRWTEBEMETAKKGLLEHGRNWSAIAIRMGVSKTVSQCKNFYFN---YKKRQN 661
Db 162 PSSTLCPICKTSLDSTSSQPNF-----NICTQCHNKVCQCGENPNPH 205
Qy 662 LDEILOQHKLKWEKERNARRKKKAPAAAEEAFAFPVVEDEEMAEASGVSGNEEWEA 721
Db 206 LTQVKEMLCNCOMQALGMDMTTAPRSKSOQLHSPALSPAHSAPAKPLGKPEQ----- 260
Qy 722 EALHASGNEVPRGCSGPATVNNSSDTESIPSP-HTEAAKDTQNGPKPBPATLGADGPPP 780
Db 261 -----ERSRSPGATQSGRQAEARAT-SVPGPTQATAPPEVGRVSPQPPUSTKPSATAP 314
Qy 781 GPPTPPRRTSRAPTEPT-----PASEATGAPT-----PPPA-----PP 813
Db 315 RPPAGEAQKSATTVPSGLGAAEQTOGGLTGKLFGLGASLLTQASTLSMVQPEADTQGP 374
Qy 814 SPSAPPVVPVKEKEETAAAPVVEGEBEQKPPAAABELAVDTGKAABEVKSECTEABEG 873
Db 375 SPSKGPPIKIVSDASKAGPRPP-----GSGPGP-----G 404
Qy 874 PAKKDAEAEATAEGA-LKAEKEGGS-----GRA-----TTAKSSGAPQDSDSAT 920
Db 405 PTPGAKTEPGPTGFGSGPGALAKTGGTSPKGRADHQAASKAAAKPKTWPKERAACPL 464
Qy 921 CSADEVDEAEAGDKN-----RL-----LSRPSLLTPT-----G 949
Db 465 CQA-ELNVGSRGPANYNTCTACKLRVCTLCGFNPTPHLVEKTEWCLNCLNCKRLLEGSIG 523
Qy 950 DPR--ANASPOKLDLKLKORAAAIPIQVTKVHEPPREDAAATKPAAPPAPPQNPQ 1007
Db 524 EPAPLPLPTQEP--PAGVPORAAAGASPLK---QKGPQGPQSGSLPPKASP-----QA 573
Qy 1008 ESDAPQPGSSPRCK-----SRSPAPPADKFAFAEAQKLPQDPPCWT----- 1050
Db 574 AKASFOAAKASPOAKPLRASEPSKTSAPSEKKTGIPVKAEPVFKPPETAVPPGTPKAK 633
Qy 1051 SGLPFPVPPREVAKSHAP-----DPSAPSYAPPQGHPLPLGLHDT----- 1091
Db 634 SGVKTDPATPVKVPVEAPKSGEABEPVVKPYQDLRSQSLSLDTGYSSDGVSSSOSE 693
Qy 1092 -----ARVLPRPPTTISNPPPLISSAKHPSVLIERQIGAISQGMVSLHVPY 1137
Db 694 ITGVVQVEVEQLDSAGVTGPRPS-----PSELHKVSSMRPSLEAQAVAPS 740
Qy 1138 SEHAKAPVGPVTMGLPLPMDPKLAPFGVKQEQLSRPGQAGPPES-----LGVPTAQE 1191
Db 741 GEWSKPPSG-----SAVEDQRRPHSLSIMPEAFDSDBELGDLBED 782
Qy 1192 ASVLRG-----TALGSVPGSGITK-----GI-----PSTRV----- 1217
Db 783 DSLANGQREQDQTAESDDFGSLRHDYVEDSSEGLSLPLPPOPPARADMTDEEFMRQ 842
Qy 1218 -----PSDSAITVYRGSITHTGTPADVLYKGTITRIIGEDSPSRDLGRGD-SLP 1264
Db 843 ILEMSAEDNLEEDDTAVSGRLAKHGAQK-----ASARPPRESSQSSVALP 889
Qy 1265 KGHVYEGKKGH---VLSYEGGSMVTCQSKEDG----- 1294
Db 890 KRRLPHNATTGYEELLSEEGPAEPTDQALQGLRRFKTIGLNSTGRLWSTSLDLQGGSDP 949

```
QY 1295 -----RSSGGPP-----HETAAKRTYDMMEGRVGRAISSASIEGLMGRAP 1337
Db 950 NLDREPELEWESLTGSPEDSRGHSSTLPASTPSYTSGTSPISLSLEED---SDSSPS 1006
QY 1338 ERHSPHLKBO----HHIRGSIQTGIPRSVVEAQEDYLRREAKLLKREGTTPPPPSRDL 1393
Db 1007 RQRLEBAKQOKARHSHGFL--LPTIEDSSEELREBEELLREQEKWREVEQOQIR 1063
QY 1394 TEAYKTO-----ALGPLKPAHEGLVATVKEAGRSIHETPRELHRTPE 1438
Db 1064 STARKTRDKELRAQRRRRSKTPPSNLSP-----IEDAS-----PTELROAE 1109
QY 1439 LPLAPRLKESITQTGTPKYDTGASTTGKCHDVRLSIGPGRTFPVHPHPLDVMADARA 1498
Db 1110 M-----EELHSSSEYSP-----SPSLDSEAEITLDCGPTFLYKSGSEYNLPAM-- 1154
QY 1499 LERACYBESLSRPGTASSGGSIARGAPVIVPELG---KPROSPLYEDHGAPFAGHLP 1555
Db 1155 -----SLCSPTETPSSGSTTTPSSGRPLKSAEAYEDWMKAEELLQROGQOAGARGP 1206
QY 1556 RGSPTVWRETPRLQEGSL-----SSKASQDKRLTSTPREIAKSPHSTVPEHHPH 1606
Db 1207 HGGP---SQPTGRSQGSFEYQDTLDHDYGRASQP--AADGTGPAFLG---ATV----- 1252
QY 1607 PISPYEHLRGVSGVDLYR--SHIPLAFDPTSIPIRGIPLDAAAYILPHRLAPNPTYPHL 1664
Db 1253 -----YELIQTOSQIAWRQASSRDLAFTEDKKKEKQFLNAESAYMDPMKONGGLTPGT 1308
QY 1665 YPYLIR--GYP-DTAALENROTIIINDYITSQQMHNTATAMAQRADMLRGLSPRESSLA 1721
Db 1309 SPTGLAAPVGFPTSTSDSGGRVIPDVRVTQ----- 1340
QY 1722 LNYAAGRGIIDLQVPHLPVLPPT-----PGTPATAMDELAVLPTAPQPFSSRHS 1773
Db 1341 -HFAKEPQEPKLHSSPASPLASKEYGWTFSQGGTGPAT-----TAMAPCPASLP 1390
QY 1774 SSPLSPGCPHLTKPTTTSSSERDRDRDREREKSLTSTTTTVEHAPTWRPQTEQ 1833
Db 1391 RGYWTPAGER--SPSTSSITHSGQPTTANYGSQTEE-----LPHAPSGPAGSR 1440
QY 1834 SSGSSGSGSGGG-----GSSS--RPASHSHAHQHP--ISPR-TQDALQ 1873
Db 1441 ASREKPLSGGDEGVGPQPSRGYSYFTGSSPPLSPSTPSESPTSPSKLGRATAEFTQ 1500
QY 1874 RPSVLHNTGKGIITAVEPSKPTVLRSTSSVVRPAATPPATHCPGLGTLGDGVYITLM 1933
Db 1501 TPS-----LTPSS-DIPRSVGTSPMVAQGTQTP--HRPSTPLVWQQSQE 1544
QY 1934 EPVL---LPKEAPRVARPERPRADT-----GHAF-LAKPPARGSL--EPASSP 1975
Db 1545 APVNVITLADASSQTRMWHASASTSLCSPTDSQPASHYSQTTTPSASQMPSEPAGPP 1604
QY 1976 SKGSEPRPLVPVSGHATI---ARTPAKNLA-----PHIASDPD-PAPPASADPHRE 2024
Db 1605 ---GFPRAPAGVDPGLALYGWALPAENISLCRISSVPGTSRVEPGPRPGTAVVDLRT 1661
QY 2025 KTQSKPFSI--QELELSLGHGSSYSPEGVFVPSVSSSLTHDXGLPHLELDKSH- 2081
Db 1662 AVKPTPIILTDQGMDLTSLAVEARKYG-----LALDPVPGRQSTAVQPLVNLNAQEQHTT 1717
QY 2082 -----LEGLRKPQGPVKLGGEEAAHPLRLPSPESQPSSSL----- 2119
Db 1718 FLATATTVISIMASSVLMAQKQKQPVVVD-----PFQSRDLDFGSGSGSPVCLAQVKVEQ 1772
QY 2120 -LQAP---GVKHORVVVTLAQH-----ISEVITQDYTRHHQQLSAPLPLAPLY 2164
Db 1773 AVQTAPYRGGRGRPREAKFARYMLPNQVTPPLARRDILITOMGTAQSVSLKPGSVPE--- 1829
QY 2165 SFGCASCPVLDLRPPSDLYLPPDHGAPARGSPHSGGKSPSPNKTSVLGGCE----- 2219
Db 1830 --PGAE-----PHRATPRLRAHALPGRKPH---TVVQMGEGAAGT 1867
QY 2220 -----DGIEPVS-----PPEGMTE---P 2234
```

```
Db 1868 VTLLPEEPACALDLTGMRPESRLACCDMAYKFPFGSSCTGTETHPADSPDKSVTDAALP 1927
QY 2235 CHRSVAVPLYLYDGEQTEP---SRMGSKSPGNTSQ-----PPAFFSKLTESN--SA 2281
Db 1928 GQSSGPFYS--PRDPPPEPPLTFRAQGVWVGPHPEORPYPOGLFGLRYSMSDTNLAEA 1985
QY 2282 MVKSKQEIKNKLNTHRN----- 2306
Db 1986 GLNYHAQRICQLFGQGRDSDAVLSSLKHSYSLGFADGRVYLGQLOQYGSFTDLRHPTDLL 2045
QY 2307 SQP-----GTEIFNMPAIT----- 2320
Db 2046 SHPLPMRPYSSVNIYSDHRYGRGDAVGFOEASLAQYSATTAREISRMCAALNSMDQYG 2105
QY 2321 -----GTGLMTRYSOAQVEHA---STWGLEAIIRKALMGK--YDWEESPLSANA 2367
Db 2106 GRHGGSGGPDLYVPOQ---HGPGLNAPQGL-ASLRSGLLGNPTYPEGQPSFNLAQ- 2159
QY 2368 FNPLNASAS-----LPAAMPITAADGRSDHLLTSP----- 2397
Db 2160 YGPAASOGTAVROLLPSTATVRAADGMISYINTPIAATLPIITQPASVLRPMVRGGMRY 2219
QY 2398 --GGGKAKVSGRPSRKAQSP--APGLASG-----DRPPSVSVHSEGDENRT 2443
Db 2220 PYGSGG---VTAVPLTSLTRVPMIAPRVPLGPAGLYRVPAPSRFFIASTI--- 2267
QY 2444 PLNRWVEDRP-----SSAGSTPFP-----YNPLMFLQAGVMASSPPPG 2483
Db 2268 PASEGVYVLGPPAAKASGAGGPPRPELPAGAREEPLSTTAPPVAVIKEAPVQAQAPAPP 2327
QY 2484 -LPA-----GSGPLAGP 2494
Db 2328 QKPAGDAAGSGSGVLGRP 2346
```

RESULT 45
Q7XXN1
ID Q7XXN1 PRELIMINARY; PRT; 4322 AA.
AC Q7XXN1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein p0705E11.8.
GN P0705E11.8.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Oliveira A.C., Mattos L.T., Carvalho F.F., Shimano A., Zimmer P.D.,
RA Malone G., Dellagostin O.;
RT "Oryza sativa nipponbare genomic DNA, chromosome 9, PAC
clone.P0705E11";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP006548; BAC79158.1; -;
KW Hypothetical protein.
SQ SEQUENCE 4322 AA; 469343 MW; E5A146B2FIDE5A75 CRC64;

Query Match 3.78; Score 493.5; DB 10; Length 4322;
Best Local Similarity 19.38; Pred. No. 7.7e-15;
Matches 570; Conservative 333; Mismatches 1098; Indels 947; Gaps 137;
QY 4 STQLVAQTRATEP-----RYP--PHLSYYPVQIARTHTDVGLLEYQHHSRDYASHLSPGS 57
Db 1181 SKQVIADKALSVTLPPTRYPTAEDTSRQPKRTGSTPTERAVQDGRDAFRESKSVDSSTSS 1240
QY 58 IIQPOQ--RRRPSLSSEFPQGNERSQELHLRPESHSLPELCKSEMEFESKPRLELLPD 115
Db 1241 TQPSGYRRAAASLPKQEAADSRSTPL---PFQRRY-PDVEHT-----TKEP----- 1283


```
QY 1812 KSILTSITTVHAPIWRRTGTEQSSGSSGSGGSSSRPASHAHQHSPISRTQDAL 1871
Db 2332 -----VTYRSSLPLEPSD-----PVMSKKEPEL 2356
QY 1872 QORPSVLHNTGMGLITAVBPSKPTVLKSTSTSSPVRPAATFPFATHCPLGTLGDVYPT 1931
Db 2357 TQE-----EIDHTARIQSLAQSFEQASTVPDRPPLPVR-----LPT 2394
QY 1932 LMEVLLPKE-----APRVARPERPRADTGHAFKAPPARSGLEPASSPSKSEPRPL 1984
Db 2395 VDEPVSSEQHEDRRSSATSGADYERF--FDQFTYERSPL--LEPCEEVPVMSKKEPEL 2450
QY 1985 VPPVSGHATARTAKNALPHAGDPDPAPASADPHREKTQSKPF--SIQELRLSLG 2042
Db 2451 TQEBIDHIAWTSIAEQSSFEQASTIPDRPPL-----PVRLPTVDEPVSSEQHEDRSSA 2506
QY 2043 YHGSSYSEGEVPEVPSVSSPSLTHTDKGLPKHLELDKSHLGELRPQPGVKLGGEAAH 2102
Db 2507 TSGADYERSFDQV-----TYERSPL--LEPYEFPVMSKV-----BELTQEEIDH 2551
QY 2103 LPHLRPLPESQ-----PSSPILQTAQGVKGQRVVTIAQH-----ISEVITQDYTRH 2150
Db 2552 IARIQSLAQSFEQASTIPDRPPLPVRLPTV--DEPIVSEQHEDRRSSATSGADYERS 2609
QY 2151 HPQI-----SAPL--PAPLYSFGASCVPDLRPPSDLYLPDPDHGAPARGSPHSEGGK 2204
Db 2610 FDQDVTYERSPLIPTPSSSFEQAS--TVPD--RPPLPVRLPTVDE--PIVSS----- 2657
QY 2205 RSPENKTSVLGGEDGIEPVSPGEMTEPGHRSAYV-----PLLYRDEGQTEPSRMGSK 2260
Db 2658 EQHEDRRSSATSGAD-----YERFDQDVTYERSPL--BPCBEPVMSK 2701
QY 2261 SPGNITSOP--PAFFSKLITESNS-----AMVSKKQBIQINKLN 2295
Db 2702 EPELTQEEIDHIAWTSIAEQSSFEQASTVPDRPPLPVRLPTVDEPVSSEQHEDRRSSA 2761
QY 2296 THNRN-----EPEYNISQGTETFMNPAITGTLTYRSQAOVQOEHASTNMGLEAIRK 2348
Db 2762 TSGADYERSFDQDVTYERSPLPCEBPV-----MESKEPELTQEEIDHIAWTSIAEQ 2816
QY 2349 ALMKYDOWESSPLSANAFNPLNASASLPAA--MPITAADGRSDHTLSPCG 2399
Db 2817 SSFEQASTIPDRPPLPVR-----LPTVDEPVSSEQHEDRRSSATSG 2858

RESULT 48
Q61479
ID Q61479 PRELIMINARY; PRT; 2282 AA.
AC Q61479;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA binding protein Rc.
GN KRC OR Rc.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Thymocytes;
RA MEDLINE=97001141; PubMed=8812474;
RA Wu L.C., Liu Y., Strandmann J., Mak C.H., Lee B., Li Z., Yu C.Y.;
RT "The mouse DNA binding protein Rc for the kappa B motif of
RT transcription and for the V(D)J recombination signal sequences
RT contains composite DNA-protein interaction domains and belongs to a
RT new family of large transcriptional proteins.";
RL Genomics 35:415-424 (1996).
DR EMBL; L46815; AAB40884.1; --
DR PIR; T42717; T42717.
DR HSP; P15822; IBB0.
DR MGD; MGI:106589; Krc.
DR GO; GO:0005634; C:nucleus; IDA.
```

```
DR GO; GO:0003677; F:DNA binding; IDA.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf_C2H2; 5.
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00151; ZINC_FINGER_C2H2_2; 5.
KW DNA-binding; Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 2282 AA; 246919 MW; E4404BBC8CF7B2C CRC64;

Query Match 3.6%; Score 477; DB 11; Length 2282;
Best Local Similarity 19.3%; Pred. No. 2.2e-14;
Matches 518; Conservative 288; Mismatches 900; Indels 976; Gaps 127;

QY 7 LVAQTWRATEPRYPHSLSYVQIARTHTDVLGLEYQHSDYASHLSPGSIIOQRERP 66
Db 107 LEGSTWOLVDMRPGSGSFVA--PGSHQSQQL-----PSHAS-----ILPPELP 151
QY 67 SLLSEFOPGHERSQELHLRPESHSHYLPGLKSEMEFIESKRP-----LELLDPLLRP 120
Db 152 GIPKVFVP--RPSQVSLKPAEAAHKE-----RKQKPGKYICQYCSRCAKP 197
QY 121 SPLLA-----TGO-----PAGSEDLTKDRSLTKGLEPVSPPSPH-----TDPELEL 162
Db 198 SVLQKHRTSHGTGERPYCGPCGCFKTKSNLYKHR-----KSHAHRIKAGLASGSEM 251
QY 163 VPPRLSKEELIQNNDRVDREITWVEQITSLKKKQQOLEEAAKPEPEKPVSPPIESK 222
Db 252 YPPGL-----EMERIPG-----EEPEEPEGESTDEETGAASGPDVLKPK--K 296
QY 223 HRSVLQIYD--ENRKAEEAAHRILEGLGPOV-----ELPLNQSDTRQYHEN 269
Db 297 HPLLSSLSYSGSHGSGQERCSLSQSTGSPLEDPAFAEASSEHPLSHKPEDTHTIKQK 356
QY 270 IKINQAMRKXLLLYFKERNHARKQKQFCQRYDQLEALEKKVRIENRRRAKESKV 329
Db 357 LALRLSERKKLI---EQTFLSPGSGKSTESGYFSRSESAEQV---SPNTNAK--- 405
QY 330 REYYEKOFPEIRKQRELQERMQSRVQSGSLNSAARSEHEVSEIIDLSEQNLEKQM 389
Db 406 -SYAEIIFGKC-----GRIGQRTSLASTSTQ----- 431
QY 390 RQLAVIPPMLYDADQQRIFKINMGLMADPMKYDKQVMMWSEQEKETREKPMQHPK 449
Db 432 -----PLPLSSEDKPSLV-----PLSVP-----RTQVIEH-- 457
QY 450 NFGLIASFLERKTVACVLYYLTKNENYK---SLVRBSYRRRGKSSQOQQOQQOQQ 505
Db 458 -----ITKLI---TINEAVD---TSEIDSVKPRRSLTRRSSVSPKSS----- 496
QY 506 QQQQQPMRSPSSQEEKKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE 550
Db 497 -----LYRDSLSSHGKTKQEQSLLSLQHPSPSTHPVPLLRSHSMPSAACTISTHHHTF 550
QY 551 -DTSGBDNDEKEAVASKRKTANSQGR---RKGRITRSMANANSBEAIPQQAELAS 606
Db 551 RGSYSFDHDVADPEVPSRNTPVFTSHRMLKRAHAIELPLGGEYSSEB---PGPSSKDPT 607
QY 607 MELNESSRWTEEM-ETAKKGLLEHGRNWSAIARMVGSKTVSQCKNFYFNKYKQNLD-- 663
Db 608 SKPSDEPEPKESDLTKTKTKGFKTKGANY-----ECTICARYKKKRDNYBAH 654
QY 664 -----EILQOHKL---KWEKER-----NARKKK----- 684
Db 655 KKYVCSELQITKAHSAVGAHEVEKTAQEPWSQMMHYKLGATLELTPLRKEKEKSLGDE 714
QY 685 -----KAPAAASEEAAFPVVEDEEMWASGVSNBEMVEEA 721
Db 715 EEPFAFACPGPSETAHRNRLGTSKSPAASKSA---PSLEDPRASSPGLPSQE----- 764
QY 722 EALHASQNEVPRGE--CSGPATV--NNSSDTESTPSPTHEAAKDTGQNGPKPATLGADGP 778
Db 765 ----LGQNQGRGEOCPKFTVIOHTSSFEKSDP-----PEQPSGLEEDKP 806
```

QY	779	P-----	PGPTPRRTSRAP	IEPTPASEATGAPT	PPPPAPSPSAPP	PPVVPKKEEETAAP	835	
Db	807	PAQFSSPP	PHGRSAHSLQ	PLVRQ-----	PNIQVEILVTEP	DRPDTEPEP	855	
QY	836	PVEEGEKQ	PPAAEELAVD	TGKAEEPVKSECTEE--	AEEGPAKGD	-----AAEAATAEGA	890	
Db	856	PKPEPEK-----	TEEQWPORTQ	TLAQLPAELKPLPKKKRLRLAEMAQ	SGESS	903		
QY	891	LKAEKKEGSGR	ATTAKSGAPO	SDSSATCSADE	VEDEAGGDKNLLSP	RP-----	942	
Db	904	FESSVP-----	LRSFSPQSSSI	SLSGSRASAF	REDHGKAE-----	APGPFSDTRSKT	952	
QY	943	-----	SLITPTGDP	-----	RANASOKP	-----LDLKOLKORAAAIPIQV	978	
Db	953	LGSHMLTV	PHHPHAREMR	RSASQSPNV	PHSHMTTRSKS	FDYGLSPTGFSL	1007	
QY	979	TKVHEPPRE	DAAPT	KPAPPAPPQ	-----	NLOPESDAPOO	-----G	1016
Db	1008	-----	AVPAAPPPPA	APPERRKCF	VLVQA	SLNRPEAELEAVPRENKRAVRMPAA	1057	
QY	1017	SSPRGKSRP	-----	APPADKEA	FAAQAOKLP	-----GDPPCWT	-----GLFPV	1057
Db	1058	SKPTKSSVP	QISVGT	TGGPGSGK	-----	SQMQRPLGSSP	PYTEALQVFP	1113
QY	1058	P-----	PREVIKAS	HAPDPS	-----	AFSYAPFGHPL	PLGLHDTARVLP	1102
Db	1114	PASLFSIQ	LLLPQOQSS	BEFFTOAMAG	LLSSYSNP	--PLP	SLFQA--PPLP	1169
QY	1103	SNP	-----	PPLISSAKH	-----	PSVILEROIGAISQMSVOL	-----HV	1135
Db	1170	LHPSQLHL	PQLPHAADIP	FPQPPSFLP	MPCPAPSTLSG	VFLPQSFALQ	POEIESHL	1229
QY	1136	PYSEHAKAP	GVPTWGLPL	MDPKKLAP	FGS	-----	VKOEQLSPRCAQBP	1189
Db	1230	P---	PVKTS	LPPLATG	---PPGSS	STEYSDDIQ	LPVPTQATS	1281
QY	1190	QEASV	-----	LRGTAL	GSV	-----	PGGSITKIGIPSTRVPS	1227
Db	1282	PDAVSLV	VPVRIQTHMPS	YCSAMVTL	LSQLVTQSG	SPASTALTKY	EEPSKSMVT	1339
QY	1228	SITHGT	PADVLKGTIT	RIIGEDSP	RDRGRED	SLPKGVHYEG	KGHVLSYEG	1287
Db	1340	-----	CEADV	YE-----	AEPG	SSISKEQNRGYQTPYL	RVPERKGTLS	1386
QY	1288	QCS	-----	KEDGR	SSGPPH	-----	-----TAAPK	1308
Db	1387	GCSSTAS	GSKRVLSPAG	SLTMTQOQ	KRVKEEASKA	DEKLELVTS	CVSVLTSTEDRK	1446
QY	1309	RTYDMEGR	VGRA	-----	ISSA	TEGLMGRAI	PPERHSPH	1361
Db	1447	KTEKPHV	CGQSRREAE	TLSSLSDVSD	PKELSP	LSHS	-----	1494
QY	1362	SYVEAQE	DYLREAKL	KREGTPPP	PPPSRDLT	EAYKTQAL	GLPKLKA	1421
Db	1495	S--	EALKE	YAQPSK	--AHR	RGGLP	-----	1520
QY	1422	GRSHEI	PRELRHTPE	LAPRLKEG	SIQTGTP	LKYDTGAST	CGSKH	1481
Db	1521	-----	EDPKE	QDLPPLAPP	SSLPUSD	---TSP	RAKLQEG	1559
QY	1482	RTFPVH	-----	PLDYM	--ADAR	ALERA--	CYEESLKS	1524
Db	1560	LQFP	SLHTTNV	SMCYLNYI	KENHIQHADR	SSVYAGWC	ISLYNPNL	1619
QY	1525	GAPVIV	PELGRPQ	PLTYEDHGA	FPAGHLP	RGSPVTVRE	PPRLOE	1584
Db	1620	-----	SKQV	SKETYTWAT	AP---	HPEA	GRVLPNSR	1667
QY	1585	LTSTPRE	IAKSPHS	-----	TVPEH	PHPTSPV	HELLRGV	1632
Db	1668	-----	DPA	VEKEKQK	AREGPT	SKRGEP	ARKIFE	1715
QY	1633	-----	DPTS	IPRGIFL	DAAYL	PRHL	-----	1672

Db	1716	GRGRYVCE	GIRCKP	SMKKH	IRTH	TVR	PPYCKCH	FAFKTKGNLT	KHMKSKAH	1732
QY	1673	YPTAAL	NRQTI	INDYIT	SQOMH	HTATAMA	QADMLR	GLSPRESS	LALNYAAG	1804
Db	1776	COET	GVLE	BELEA	-----	BEQTS	DDHLQD	-----	SEGQ	1792
QY	1733	DLSQVPH	LVLV	PPTPG	TATMD	RLAYL	PTAP	OFFSR	SSSSPLSPG	1819
Db	1805	-----	AEAVE	-----	EHQ	FS	DL	-----	DS	1852
QY	1793	SSER	DRDRDR	DRREK	SIL	TTTVE	HAP	WRP	CTE	1864
Db	1820	DS	DL	DE	EE	EE	EE	EE	EE	1911
QY	1853	SHSHAHQ	HSPT	SP	--R	TODAL	QOR	PSV	LHNT	1903
Db	1865	-----	VGP	QAP	PD	ST	DE	VE	GS	1971
QY	1912	TPT	PTH	CP	LG	TL	DV	YTT	LM	1942
Db	1904	GL	PL	GL	LA	PL	-----	E	K	2036
QY	1972	A	-----	SSP	SG	SE	PL	VP	VS	2086
Db	1943	SV	TR	KH	SL	TK	ND	SS	Q	2036
QY	2027	QSK	PF	SI	O	E	L	R	S	2128
Db	1998	RL	TP	YPI	-----	GRE	A	P	A	2128
QY	2087	RP	KQ	-----	PG	V	K	L	G	2128
Db	2037	PP	Q	T	V	L	G	K	W	2128
QY	2140	SE	V	I	T	Q	D	V	T	2128
Db	2085	P	E	T	C	S	A	M	R	2128
QY	2199	-----	H	S	E	G	K	R	S	2128
Db	2129	N	L	F	S	H	L	P	H	2128
QY	2239	S	A	V	P	L	L	Y	D	2128
Db	2187	T	G	A	-----	R	E	A	Q	2128
RESULT 49										
ID	O43161	PRELIMINARY;	PRT;	3851	AA.					
AC	O43161;									
DT	01-JUN-1998	(TEMBLrel. 06, Created)								
DT	01-AUG-1999	(TEMBLrel. 11, Last sequence update)								
DT	01-OCT-2003	(TEMBLrel. 25, Last annotation update)								
DE	BASSON	protein (Hypothetical protein KIAA0434) (Fragment).								
GN	BSN OR KIAA0434.									
OS	Homo sapiens (human)									
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
OC	Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.									
OX	NCBI_TaxID=9606;									
RN	[1]									
RP	SEQUENCE FROM N.A.									
RA	Winter C., Dieck S. T., Bockmann J., Boeckers T., Kaempf U.,									
RA	Sanmarti-Vila L., Laingnaese K., Altrock W., Stumm M., Soyke A.,									
RA	Weacker P., Garner C.C., Gundelfinger E.D.;									
RT	"The presynaptic cytonatic protein Bassoon: Sequence and chromosomal									
RT	localization of the human BSN gene."									
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.									
RN	[2]									
RP	SEQUENCE OF 2281-3851 FROM N.A.									
RC	TISSUE=Brain; Negase T., Nakajima D., Seki N., Ohira M., Miyajima N.,									
RA	Ishikawa K.,									

Db	1716	GRGRVYVCEGIRCKKPSMLKKHIRTHTRDVRVPVCKCHFAFKTKGNLTKHMKSKAHSK	1775
Qy	1673	YDPTAALNROTIINDYITTSQOMHNTATAMAQADMLRGLSPRESSLAINVAAGPRGII	1732
Db	1776	COETGVLELEA---BEGTSDDLHQD-----SEGOEG--	1804
Qy	1733	DLSQVPHLPVLVPTTGTPTATMDRLAYLPTAQPFSSRHSSPLSPGPGTHLTKPTTTS	1792
Db	1805	-----AEAVE-----BHQFSDLE-----DS	1819
Qy	1793	SSERDRDRDRDRDREREKSILTTTVEHAPIWRPGTEQSSSGSGSGGGGSSSRPA	1852
Db	1820	DSDSLDEDEEEEEEESQDELSGPCS-EAAPPCLPPTLQENSSP-----	1864
Qy	1853	SHSHAHQHPISP-RTQDALQORPVSVLHNTGMKGIIITAVEPSKPTVLRGSTSSPVRPAA	1911
Db	1865	-----VEGQAPDSTSDVPPGSSI-----SEATHLTASSCSTPSRGTO	1903
Qy	1912	TPPPATHCPGGTLDGVYPTLMEPVLVLLPKEARVARPERPRADTHAFLAKPPARSGLEP	1971
Db	1904	GLPRLGLAPL-----EKDSSAPSPKATSPRRPWS-----PSKEAGSRP	1942
Qy	1972	A-----SSPSKSGSEPRLPVPPVSGHATIAARTAKNLAPHASDPDPAPPASADPHREKT	2026
Db	1943	SVTRKXSLTKNDSSPQOCSPAREAQASVTSTPGQMGPR-----DLGPHLCGSPRELS	1997
Qy	2027	QSKPFSIOELERLSLRGVSGSSYSPGVEPVSPVSPSLTHDKLGLPHLEELDKSHLEGEL	2086
Db	1998	RLTTPYPI-----GREAPAGLERAT-----DTGTPRYSPTRRSLSLGOAES	2036
Qy	2087	RPKQ--PGPVKLGEEAAHLPH-----LRPLPESQSPSSPLLQTAGVKGHORVVTLAQHI	2139
Db	2037	PPQTVLPKQWALAGPCSPSADKSLGLGVPVRA-----LLQVPV--LPH-----TLLSR	2084
Qy	2140	SEVITQDYTRHHIPOOLSAPLPALPYSPFGASCVPVLDLRRPPSLDLYLPPPDHG-APARGSP	2198
Db	2085	PETCTSAWRKTESRSPSAG--PGELFPRPFA-----PHDF-----HGHLPSRSEE	2128
Qy	2199	-----HSEGKRSPENKTSVLGG-----GEDGIEPVSP-----EGMTEPGHSR	2238
Db	2129	NLFSLHPLHLSQLLSRAPCP--LPIPIGQIMQVQARPGAQPTVLPGPCAAWVSGFGSGGSDL	2186
Qy	2239	SAVYPLLYRDEGTEPSRMG-SKSPGNTSOPPAFFSKLTESN	2279
Db	2187	TGA-----REAQSRWSPTSPSASVSFVAKVSKFTLSS	2221
RESULT 49			
O43161			
ID	O43161	PRELIMINARY; PRT; 3851 AA.	
AC	O43161;		
DT	01-JUN-1998 (T-EMBLrel. 06, Created)		
DT	01-AUG-1999 (T-EMBLrel. 11, Last sequence update)		
DT	01-OCT-2003 (T-EMBLrel. 25, Last annotation update)		
DE	BASSON protein (Hypothetical protein KIAA0434) (Fragment).		
GN	BSN OR KIAA0434.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Winter C., Dieck S.T., Bockmann J., Boeckers T., Kaempf U.,		
RA	Sanmarti-Vila L., Langnaese K., Altmock W., Stumm M., Soyke A.,		
RA	Weickner P., Garner C.C., Gundelfinger E.D.;		
RT	"The presynaptic cytomatrix protein Bassoon: Sequence and chromosomal		
RT	localization of the human BSN gene."		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE OF 2281-3851 FROM N.A.		
RT	TISSUE=Brain;		
RC	Ishikawa K., Negase T., Nakajima D., Seki N., Ohira M., Miyajima N.,		
RA			

RESULT 49

O43161	PRELIMINARY;	PRT;	3851	AA.
ID	O43161			
AC	O43161			
DT	01-JUN-1998	(TrEMBLrel. 06, Created)		
DT	01-AUG-1999	(TrEMBLrel. 11, Last sequence update)		
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)		
DE	BASSOON	protein (Hypothetical protein KIAA0434) (Fragment).		
GN	BSN	OR KIAA0434.		
OS	Homo sapiens	(Human).		
OC	Eukaryota; Metazoa; Chordata;	Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates;	Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Winter C., Dieck S.T., Bockmann J., Boeckers T., Kaempf U.,			
RA	Samarti-Vila L., Langnaese K., Altmann W., Stumm M., Soyke A.,			
RA	Wiesacker P., Garner C.C., Gundelfinger E.D.;			
RT	"The presynaptic cytomatrix protein Bassoon: Sequence and chromosomal			
RT	localization of the human BSN gene."			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 2281-3851 FROM N.A.			
PC	TISSUE=Brain;			
RA	Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,			

QY 1724 YAAGPRGIIDLSQVPHLPVLVPTTCTPATAMDRLAYLPTAQPFSSRRHSSPLSPGGPT 1783
DB 1654 YAS-----VVIIDHPEVTVIDP-----QSNLNDGPT 1682
QY 1784 HLTKTPTTSSRRDRDRDREREKSILTSITTTVEHAPIWRGTQSSGSSGG 1843
DB 1693 EV-----VSKQKRLQDEER-----QVNNKUANEXGRS----- 1720
QY 1844 GGGSSRPASHAHQHSPISPR-----TQDALQORPSPVHNTGKGIITAVPSPKPTVL 1899
DB 1721 -----QTSKLPREFAKQATGLOAQS-----SASVPLASAPLPPTSASV 1762
QY 1900 STSTSPVRPAATFPATHCPGLGTLGVYP-TLMEPVLLPKAARVARPERPRADTGA 1958
DB 1763 PASTSAPL-----PATLTPVPASTADVPASTLAPVLASTAPVPASPLAPVASASV 1815
QY 1959 FLAKPAPS-CLEPASPSKSEPRPLVPPVSGHATI-----ATPAKN 2001
DB 1816 SASVFASTSAAAITSSAPASAPATPILASVSTPASVTILASAPILASALASTSAPT 1875
QY 2002 LAPHASPDPP---APPASADPHREKTKOSKPFSTOELERLSLGVHSGSYSPGEPVSP 2058
DB 1876 PAPAASSPAAPVITAPTIPASAP-----TASVPLA-----PASASAPAP-TP 1918
QY 2059 VSSPSLTHDKGLPKHLELDKSHLEGLRKPQGPVKGGEAAHPLRLPLPESQPSGS- 2117
DB 1919 VSAPN-----PAPPAPAQTOAQ-THKPVQNPQLTTSQSSKQ 1953
QY 2118 -----PLLOTAPGVGHQVRVVTIAHISEVITODYTRHHQO---LSAPL-----P 2160
DB 1954 PPPSIRLPSAQTPNG-----TDVASKSITQTPQSHGTLTAEIWMNDKNVAPP 1999
QY 2161 APLYSFPGASCPLDLRPPPSDLXLPDPDHGAPAR-----GSPHSEKRSPEPNKTSV 2214
DB 2000 AVLNDISKLGPI-----SPPO-----PPSVSANNPLTSFGAPSEGAKNQOE-----SG 2046
QY 2215 LGGGEGIEPVSPPEGMTEPGHRSVAVPLLYRDGEQ---TEPSRMGSKSPGNTSOPPAFF 2272
DB 2047 LEIGDTTIOGAPASNGE-----NEVVPVLEKSADKIPEPEQORQOP---RAGPIKA 2098
QY 2273 SKLTSNSAMVSKQO-EINKKLANTHNN-----EPYNN----- 2305
DB 2099 QKLPDLSPVENKEHKEPGPIGKERSLKNRVKDAQOQVEPEQOQKSPATVRSTDPVTTKET 2158
QY 2306 ---ISQPGTEIFNMALITGGLMTYRSQAVQEHASNMGLEAITRKALMGKYDQWESPL 2363
DB 2159 KAVSEMSIEGTMISVSAEYGNKAKESVDTITPSSLPNTVATNTKMTDLVNNVPL 2218
QY 2364 SANAFNP-----LNASASLPAAMPITAADGRSDHTLTSPGGGKAKVSGRPSRKAQSP-- 2417
DB 2219 PNTLPLPKRETIQSSSLTSVPPTT-----FSLTFMESARKAWENSPNVRKGSPT 2271
QY 2418 -----APGLASGDRPPSVSVHSGDCNR-RTPLTNRWEDRPSAGSTPPPNPLMR 2470
DB 2272 STAPPIATGVSSASGSPSTANTNSFSSAMPQIPVASVTPTASLSCAGT-----YTTSSL 2327
QY 2471 LQAGVMSPP-----PPCLPAGSGLA 2492
DB 2328 TKSTTTSDPPNICKVKPQOOLQTSLSLPSA 2355

RESULT 54
QSC0A3
ID Q9COA3 PRELIMINARY; PRT; 2219 AA.
AC Q9COA3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA1760 (Fragment).
GN KIAA1760.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
EX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
DR EMBL: AB051547; BAB21851.2; --
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR008271; Ser_thr_pkin_AS.
DR InterPro: IPR001245; Tyr_kinase.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 2219 AA; 234470 MW; B7AD70196C2A92C0 CRC64;
Query Match 3.5%; Score 459.5; DB 4; Length 2219;
Best Local Similarity 20.1%; Pred. No. 1.5e-13;
Matches 564; Conservative 294; Mismatches 961; Indels 985; Gaps 146;
QY 55 PGSIIPOR-RRPSLLSE-----FQGNERSQELHLRPEHSYLPGLGKSEMFIESKRP 108
DB 6 PGTLMPEGRGAGPAGMAEPRAKARPQRF-----LRSSVVESDQEEPP 50
QY 109 RLELL-----PDPLL-----PSPLLATQP-----AG 131
DB 51 GLEAAEPGPQPQPLQRRVLLCKTRLIAERAGRPAAPALVAQPGAPGADAG 110
QY 132 SEDLTORSRLGKLEFVSPSPPHPTDPELELVPPRLSKEELIQNMDRVDRITWVEQQIS 191
DB 111 PEPV-----GTQEP--GPDPIAAAVETAPADGGPREAAATVRKEDEGAARKEP 161
QY 192 KLKKKQQLLE-----AAKPEPEKPVSPPTIES 221
DB 162 RTRDEPEEEDDEDLKAVALTSLDGRFLKFDIELGRSGFTVYKGLDTETWVAVMCL 221
QY 222 KHRSLVQIIIDENRKAHAHRIIEGLGPQVELPLYNQPSDTRQYHENIKINQAMRKLI 281
DB 222 QDRKLTKL--ERQRFKEA--EMLKGL-----QHPNIVRFY--DFWESSAKGRCI 266
QY 282 L-----YFKRNHARKQWKOKFCQVDQLMEAL-----EKKVER 315
DB 267 VLVTLMATSGTLKTLKFKVKKPKVLSWCR--QILKGLLFLHTRTPPIIHRDLKCDN 323
QY 316 I-ENNPRAKESK-----REYEQF---PEIRKQRELQERMQSVGQSGLSM-S 364
DB 324 IFITGPTGSKIGDLGLATLKRASFASVIQTPFMAPEMVEEHYDESVDVYAFGCMLE 383
QY 365 AARGEHEVSEIIDGLSEQENLEKQMLAV-IPPMLYD-----ADQ 404
DB 384 MATSEYPYSEC-----QNAAQIVRYKTCGKPKASFEKVHDPEIKELIGECICKKEER 436
QY 405 QRIKFINNGLMADPMKV-----YKDRQVNMMSQEKETFRKFMQHPKNFGLI 454
DB 437 YEIKDLSHAFFAEDTGVVRVLABEDHGRKSTIALRLWVEDPK-----KLKGPKDNGAI 491
QY 455 A-SP-LERKT---VAECVL---YYLTTKKNYKSLVRSYRRR-----GKSQ 494
DB 492 EFTFLEKTEPDEVAQEMIESGFPH-----ESDVKIVAKSRDRVALIQRRERIWPAIQ 546
QY 495 QQQQQQQQQQQQQQQQMP-----RSSQEEKDEKEKEAEKEBEKPEVENDKEDL 545

Db 547 PKEQDVGSPDKARGPPVPLQVQVYTHAQAGQPGPPEPE-EPEADQHLLPPTLPTSATSL 605
Qy 546 LKEKTDITSGEDNDEKAVAGKRTANSQGRKRITRSMANEAN-----SEE 594
Db 606 ASDSTFD-SGQGS-----TVYSQSSQSVMLGSLADAPSPAQCVCPPVSEB 654
Qy 595 AITPOQSAELASMELNESSRWTEEMETAKGLLEHGRNWSAIAIARMVGSKTVCKKNFYF 654
Db 655 PVLQSLSLGAY-----QOPTAAGEL-----PVGSVFAPAC-----686
Qy 655 NYKKRONLDEILOQHLKWEKERNARRKKKAPAAASEEAAFPVVEDEMEASGVSGNE 714
Db 687 -----PPSLQOH-----LPPPST---PMPTGQOQAPP---G 723
Qy 715 EEMVEEALHASGNEVPRGSCGPATVNNSSDTESTPSPTTEAAKOTGONGPKPPATLG 774
Db 705 -----LPPST---PMPTGQOQAPP---G 723
Qy 775 ADGPPGPPPTP-PRTRAPTEP-----TPASEATGAP-----806
Db 724 QOPPLAQOTLPOVLAPQVVPVLPQVPPHPLPYLAPASQ-VGAPAKLPLQMPQAPLOP 782
Qy 807 ---TPPPAPPSPAP-----PPVVPKEKEEETAAAPPVEGE-----841
Db 783 LAQVPPQMPPIPVVPPITPLAGIDGLPPALP-----DLPTATVPVPPPPQYFSPAVILPSL 838
Qy 842 -EQKPPAAEELAVDTGKAE-----PVKSECTEEAEBAEGPAKDAEABATAEAGALKAEKK 896
Db 839 AAPLPPASPALPQAVKLPHPPGAPLAMPCTRTVFNAP-----876
Qy 897 EGGSGRAATTAKSSGAPOSDSATCSADEVDEAEAGDKNRLSPRSLTTPGDP-----951
Db 877 ---AIIPLAVAPGVAALSIHSA---VAQLPGQVYPAAPFO---MATIDVPPSPHH 925
Qy 952 ---RANASPKQLDLKQLKORAAALPPIQVTKVHEPPREDAAPTKAPAPPQPPQNTLOPE 1008
Db 926 TVQNMRATPPQ-----ALPP-----QPTLPPQVPLPPOPTLPPQ 960
Qy 1009 SDAPQOQSSPRGKRSRAPPADKAEFAAEQAQLPGDPDPCWTSGLPVPVPPREVILKASPH 1068
Db 961 PVLPPQ-----TRPPQVPLPPOQMLPPOVLPPOQ-----ALPVRPEPL---QPH 1003
Qy 1069 APDPSAFYAPP-----GHPLPLGLHDTAR-PVLPFRPP-TISNPPPLISSAKHPSVLER 1120
Db 1004 LPEQAAPATPGSQILLGHAPYAVDVAQVTVVPVPAAVLSPEPLPEVLPAPAPELLPO 1063
Qy 1121 QIGAIQ-GMSVQLHVPYSEHAKAPVGPVTWGLPLPMDPKKLAP-----FSGVKQROL 1172
Db 1064 FPSSLATVSASVQSVPTQTATLLPPANP-----PLPGGFGIASPCTVOLTVEPVQEBQA 1118
Qy 1173 SPRQAGPPESLGVPTAQEASVLRGTALGVSPPGSIITKGIPTRVPSDSAT-TYRGSITH 1231
Db 1119 S-----QDKPP-----GLPQSC-----SYGSDVTSG-----XELSDSCGAFGGGRLE 1158
Qy 1232 GTPADVLYGTITRIIGEDSPRLDRGREDLSLPKHGVIYEGKHVLSYEGGMSVTOCSK 1291
Db 1159 GRAAKHRRS-TR-----ANSRQERASRPLTILNVCNTGDK-----WVECOL 1201
Qy 1292 EDGRSSGPPHETAAKRTYDMEGRVGRAISSASIEGLMGRAIPPERHSPHLLKEOHHI 1351
Db 1202 E-----THNKHMTFKFD-LDGDAPDEIATVMVE-----1229
Qy 1352 RGSITQIGIPRSVVEAQEDYLREAKLAKRE-----GTPPPPPPSRDL---TEAVKT 1399
Db 1230 HDFILOARETTEIQKQVMDKAEADMLSEDTADRGSDPGTSPPHLSCTGLGTGEESKQS 1289
Qy 1400 QALGPL-KLKPAAHEG-----LVATVKEAGRSIHEIPREELRHTPELAPRPLKEGSIQOG 1454
Db 1290 QANAPVYQONVLHTKRWFIICPVAE-----HPAP-EAPESPPLPLSLP-PEAS-QDS 1341
Qy 1455 TPLKYDTGASTTGS---KKHDSVRLIG---SPGRTFPPVHPLDVMADALERALACVYESLK 1509

Db 1342 APYKQDLSSEKQPSFLASQQLLSQAGSNPFGAPPAPLAPSSPPVTALPQDGAAPATSTM 1401
Qy 1510 SRP-GTASSSGSGIARGAPV-IVPELGKPRQSPLTVEDHGAPFAGH-LPRG-SPVTMRE 1564
Db 1402 PEPASGASQAGP---GTQGLTSEL---ETSQPLA-ETHEAPLAVOPLVVLGAPCT---1452
Qy 1565 PTPRLOEGSSSSKASQDRKLTSTPREIAKSPHSTVPE---HHPHPIPSYEHLLRGVSGV 1621
Db 1453 PAPE---AASSTRDASAPREPLPPAP-EPSPHSGTQPQALGQAPLPLP---AAGVAV 1502
Qy 1622 DLYRSHIPLAFDPTSIPIRGIPLDAAAAYILPRHLPNPTPHLYPPYLIR-----1671
Db 1503 SLATSQLP-----SPPLG-----PTVPP-OPPSALESDEGSPPRV 1537
Qy 1672 GYDPTA-ALEN-RQTIINDYITSQOHHNTATAMAQR---ADMLRGLSPRESSLALNY 1724
Db 1538 GFVDSITKSLDEKLTULYQYHVPTSSASAGTPVEVGDRDFTLEPLRGDQPR-----1589
Qy 1725 AAGPRGIIDLSQVPHLEFVLPPTPGTATMDRLAYLPTAPQPFSSRHSSPLSPGGPTH 1784
Db 1590 -----SEVCGGDLALPPVPKAEVSGRVQL-----POPLVEK---SELAPTRGAV 1630
Qy 1785 LTKPTTSSSERERDRR---ERDREREKSIILSTTTTVEHADIPRPG-----TEQSSGS 1837
Db 1631 MEQGTSSSMTAESPRSMGLGYDRGQVASDSHVVPSPQDVPFAPVAPRVEPTDRDGE 1690
Qy 1838 SGSS-----GGGGSSSRPASHSHAHQHSPISPRTQDALQORPSVLHNTGMKGIIT 1888
Db 1691 AGESSABPPPSDMGTVGQASHPOT-----LGARALGSPKRP-----1728
Qy 1889 AVEPSKETVLIRSTSTSPFVRPAATFPFATHCPLGGTLDGVYPTLMEVPLLPKEA-PRVAR 1947
Db 1729 -----EQQDVSSPAKTVGRFSVVS-----TQDEM---TLASPHLSRYSAPPDVL 1770
Qy 1948 PERPRATGHAFLAKPARSGL-----EPASSPKSGSEPRPLVP-----PUSGHAT--1993
Db 1771 DEAPSSPDVLAVERAQTASSIEVGVGPEVSSDGDGPRARPPVQKASLFSVSGVAGD 1830
Qy 1994 -IARTPAKNLAPHIA---SPDPPA-----PPASASDPHREKTQSKPFISOELRSLGY 2043
Db 1831 FVKATAPLORPSPRAGSLGPETPSRVGMKVTTISVTFHSGSSYISSDNSELE-----1884
Qy 2044 HGSSYSPEGVPEVPSVSPSLTHDKGLPKHLEELDKSHLE---GELRPKQPGFVKLGGEAA 2101
Db 1885 -----DADIKKELQSLREXKHLKEISELOQSQKEI---EAL 1917
Qy 2102 HPLHLRPLP-----ESQPSPPILQATPGYKGHQVRVTLAQHISEV 2142
Db 1918 YRRLKGLPLPPNVGFFHTAPTPTRRRRTSKGLKAGKLLN---PLVRQLKVASSTGHLAD-1974
Qy 2143 ITQDYTRHHPPOOLSAPLPAPLYSPGASCPLDLRRP-----PSDLYLPPPDPHGAPAR 2195
Db 1975 SSRGPPAKDPAQASVGLTA---DSTGLSGKAVQTQPCSVRASLSSDICSGLASDGGAR 2031
Qy 2196 GS-----PHSE---GGKRSPEPNKTSVLGGEGDIEBPVPPPEGMTEPHGHSASVYPLLYR 2247
Db 2032 GQGTVVVHTPSRVTVYKSSSKPRARFLSG-----PVS-----VSIWSALKR 2072
Qy 2248 ---DGEQTESRMCSKSPGNTSQP-PAFFSKLTESNAMSVMKSKOEINKKLNTNRNEPE 2303
Db 2073 LCLGKHSRSTSSSLAPGPPGPQALHVOAQVYNNN---NNKGTFTDDLH-----2121
Qy 2304 YNISQPGTEIFNPAITGTGLMTVRSQAVOSHASTNMGLEAIRKALMGKYDOWEESPL 2363
Db 2122 -----KLVDWTSKTVGAQOL---KPTLNQLKQTKQLQDM 2153
Qy 2364 SANAFNPLNASLPAAMPITADGRSDHTLTSPGGGKAKVSGRPSRKAAPSAGLAS 2423
Db 2154 EAQAG-----WAAPGEA-RAMTAPRAG-----VGWP-----RLPPAPGPLS 2188
Qy 2424 GDRPPSVS---SVHSEGDCHNRPTLTNRVWEDRPSAGSTPPY 2464
Db 2189 TTVIPGAAPTLSVPTPGSCGR-----AVSTPTSY 2218

```

RESULT 55
ID Q9VH10 PRELIMINARY; PRT; 3111 AA.
AC Q9VH10;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CG3996 protein.
GN CG3996.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foder C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003686; AAF54511.2; ..
DR FlyBase; FBgn0037800; CG3996.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR000195; RabGAP_TBC.
DR Pfam; PF00566; TBC; 1.
DR SMART; SM00164; TBC; 1.
DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS00086; TBC_RABGAP; 1.
SQ SEQUENCE 3111 AA; 349635 MW; EDA9A5FD38115773 CRC64;

```

Query Match 3.5%; Score 458.5; DB 5; Length 3111;
 Best Local Similarity 18.1%; Pred. No. 2.5e-13;
 Matches 534; Conservative 376; Mismatches 1061; Indels 975; Gaps 123;

RT "protein binding of a DRPLA family through arginine-glutamic acid
 RL dipeptide repeats is enhanced by extended polyglutamine."
 Hum. Mol. Genet. 9:1433-1442(2000).
 CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 DR EMBL: AB036737; BA95898.1; -.
 DR Genew; HGNC:9965; RERE.
 DR GO: GO:0005634; C:nucleus; NAS.
 DR GO: GO:0005515; F:protein binding; NAS.
 DR InterPro: IPR002951; Atrophin.
 DR InterPro: IPR001025; BAH.
 DR InterPro: IPR000949; ELM2.
 DR InterPro: IPR001005; Myb_DNA_binding.
 DR InterPro: IPR000679; Znf_GATA.
 DR Pfam: PF01426; BAH; 1.
 DR Pfam: PF01448; ELM2; 1.
 DR Pfam: PF00320; GATA; 1.
 DR Pfam: PF00249; myb_DNA-binding; 1.
 DR SMART: SM00439; BAH; 1.
 DR SMART: SM00717; SANT; 1.
 DR SMART: SM00401; Znf_GATA; 1.
 KW DNA-binding; Nuclear protein.
 SQ SEQUENCE 1566 AA; 172336 MW; 4047F1530F198E2C CRC64;

Query Match 3.4%; Score 452.5; DB 4; Length 1566;
 Best Local Similarity 19.0%; Pred. No. 2e-13;
 Matches 403; Conservative 205; Mismatches 660; Indels 853; Gaps 90;

QY 204 AAKPEPEKVPSPPIESK-----HRSLVQIIYDENRKAEEAH 242
 DB 158 ASQPPQHLSEAGRGVSGKRDHLLMNVKYYRQSEVPDSVYQHLVQDRHNENSGREL- 216
 QY 243 RILEGLQFQVELPLYNQPSDTRQYHENIKINQAMKLLILYFKRNHARKQWKOK----- 297
 DB 217 -ITDPVKNREL-FISDYVDT--YIA-----AALGKCNISHFSDIPAAREFKARVDSFP 267
 QY 298 FCQRYDQLEALEKKVERIENPRRAKESKVREYYEKQFPEIRKQRELOERMOSRVQOR 357
 DB 268 YILGYNPETRLNLSQGEIRVGPSHQAKLPDLQPPSPDGDVTVTQHEEL-----VMMP 320
 QY 358 GSG-----LSMSARSEHEVSEIDLSEQENL-----EKQRLQAVIPWMLYDADQOR 406
 DB 321 GVNDCDLLMYLRAARSMAPAGCMDCGGTDEGCVAAASRDDTTLNALNTLHESGYDAGKA- 379
 QY 407 IKFINMGLMADPMKYVDROVMNWSQEKETFEKFMQHPKNGFLI-ASFLEKKTVAE 465
 DB 380 -----LQRLVKPV-----PKLIEKCTEDEVKRFVKGURQYQGNPFRIKELLNPKETGE 430
 QY 466 CVLYYYLTGKN-ENYKSLVRSYRRRGKSQQQQQQQQQQQQQQQQQQQQPPRSSQBEKDEKE 524
 DB 431 LITFYVYWKTPEAASSRAHRRRQAVR-----HHG-----RI 462
 QY 525 KEKEAKKEEKEPEVENDKE--DLLKBTDDTSGEDNDEKAVASKRKTANSQGRKGR 582
 DB 463 KTRTASTVNTPSRPPSEFLDLSASRDDFDSEDEQE-----LKGYA 506
 QY 593 TRSMANEANSEAITPQSOAELASNELNESSRWTEEMETAKGLLHGRNWSAIARMVG 642
 DB 507 CRHCFPTTSKD-----W-----HHG-----G 522
 QY 643 SKTVSQCKNFYFNKKRQNLDEILQOHLKMEKERNARRKKKAPAAASEEAPPPVVVD 702
 DB 523 RENILLCTDCRIHFKYGELPPI-----EKPVDP-----PPFMKPKVKEE 562
 QY 703 EEMASGVSGNEEMVEBAELAHAGNVPRGECGPGATVNNSDTESIPSPHTEAKDT 762
 DB 563 DD-----GLSGKHSMTRES---RGSMTLSRGRKKQAPASPDGRT-----SPINEDIRS 609
 QY 763 GONGKPKPATLGADGPPGPTPPRRTSRAPLETPPASEATCAPTPPAPSPAPPPV 822
 DB 610 GRNSPSAASTSSND-----SKA-----ETV 629

QY 823 PKESKBETAAPVVEGEQKPPAAEELAVDTGKAEEPVKSECTEAE-----EGP 874
 DB 630 KKSARKVKEEASPLKSNKQR-----EKVASDTEADR-TSSKTKTQESRPNSPSG- 683
 QY 875 AKGDAEAAEATAEGALKAEKKEGSGRATTAKSSGAPQDSDSSATCSADEADEEGDK 934
 DB 684 -EGSSDSRSVNDGSS--SDPKDIDQNRSTSPSPQDNESDSDSAQQ----- 731
 QY 935 NRLLSPRSLLTPTGCDPRA--NASPQKP-LDLKQLKORAAAIPTQVTKVHEPPREDAA 991
 DB 732 QMLQAPQALQAPGTVPAPSSAPGTPQLPTPGTPTSATAVPP-----QGSPTASAP 785
 QY 992 TKP-APPAPPPQNLQESDAPQPGSSPRGKSSPAPPADKFAFAEAQKLPGDPPCWT 1050
 DB 786 NQQAAPTAPVPHTHIQ-----QAPALHPQ-RPPSPHPPH----- 819
 QY 1051 SGLPPVPVPREVIKASPHADPSAFSAPPQGHPL-----PLGLHD-TARPVL--PRPTIS 1103
 DB 820 ---PSPHPPLQPLTGS--AGQPSAPSHAQ--PLHGQPGPGPHSLQAGPLLQHPGFPQPF 872
 QY 1104 NPPLLISSAKHPSVLERQIGAISQCMSVOLHVPYSEHAKAPVGPVTMGLPLMDPKKLAP 1163
 DB 873 GLPQAASQGAQ-----LGTSPAAAYPH-----SLQLPASQAL-- 907
 QY 1164 FSGVQEQQLSPRGOAGPPESLGVPTAQEASVLRTALGSGVPGSITKIGIPSTRVPSDAI 1223
 DB 908 ----SQQPPREQLPFGLMPH-----IKPPTTPIQLAA- 941
 QY 1224 TYRGSITHGTPADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHVLSYEGG 1283
 DB 942 -----POAH-----KHPPHLSGSPS 956
 QY 1284 MSVTCQSKEDGRSSGPPHETAAPKRTYDMMEGRVRAISSAIEGLMGRAIPERHSPH 1343
 DB 957 FSM-----NANLPPPAKP-----LSSLSTH-----PSAHPPP 987
 QY 1344 -HLKEQHIRGSIQTGIPRSYVEAQEDYLREAKLKRGGTPPPPPSRDTEAYKTOAL 1402
 DB 988 LQLMPQ-----SQPLSS--PAQPPGLTQSNL-----PPPPASHPT-- 1023
 QY 1403 GPLKKAHEGLVATVKEAGRSIHEIPREELRHTPELPAPPLKEGSITOGTPLKYDTG 1462
 DB 1024 -----GLHQV-----APQPPFAQHPPFVPGPPPTTPT-- 1051
 QY 1463 ASTTGSKKHDVRSLLIGSPGRTPFPVHPLDVMDARALERACYESLSKSRPGTASSGSI 1522
 DB 1052 -----CPSTSTPPAGP-----GTSQAQPPC-----SGAASGSI 1080
 QY 1523 ARGAPVIVPELGKPRQSPLTYEDHGAFLPRGSPVTWREPTPRLOEGLSSSKASQD 1582
 DB 1081 AGGS-----SCPL-----PTVQIKEALDDAEPE- 1105
 QY 1583 RKLSTPRETAKSPHSTV---PEHHPHIPSPYEHLLRGVSG---VDLYRSHIPLAFDPTS 1636
 DB 1106 ---SPPPPPRSPSPSEPTVVDTPSHASQASARFYKHLRGYNSCARTDLY--FMPLA----- 1155
 QY 1637 IPRGILDAAAAYVLRHLAPNPTYPHLYPPYLIRGVPDAALENRQTIINDYITTSQOMH 1696
 DB 1156 ----- 1155
 QY 1697 HNTATAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLVPPPTGTPATMD 1756
 DB 1156 ---GSKLAKKEEAIEKAREAE----- 1175
 QY 1757 RLAYLPTAPOFFSRHSSSPLSPGPGTHLTKPTTSSSERDRDRDRDRDREREKSILT 1816
 DB 1176 -----OKAREEREKEKEKERERE----- 1199
 QY 1817 STTTVEHAPIWRPCTEQSGSSGSGSGGSSSPASHAHQHSPISPRTQDALQORPS 1876
 DB 1200 -----REARAKAASSAHEGRSLDFQLSGGHMRPSPPEPTTTAAVPPY 1245
 QY 1877 VLHNTGMKGIITAVEPSKPTVLRSTSTSSPVRPAATPPPATHCPLGGT-----LDGVY 1929

```
Db 1246 IGPDT--PALRTLSEYARPHVMSPTNRNHPF-----YMLNPTDPLLAHYHMPGLY 1293
Qy 1930 ---PTLMEPVLLPKF-APRVARPERPRADTGHAFKAPKARSGLBPASSPSKSGSEPRPLV 1985
Db 1294 NVDTIRERELREIREIREIRERELRMKPGFEVXPPELDPJLHPAAN-----1342
Qy 1986 PPVSCHATIARTPAKNLAPHASDPDPAPPASADPHREKTQSKPFS-----IOELRLS 2040
Db 1343 -----PMEHFARHSALITPP-----TAGPH-----PFASFHPLGNPLERER 1378
Qy 2041 LGYHGSSYSPE-----GVPEFVSFVSPTSITHD-----KGLPKHLEBLDKSHLEGE 2085
Db 1379 LALAGPQLRPMSYDRLAERIAERASWLSLSDPLARLQMFNVTPHHQH---SHISH 1435
Qy 2086 LRPKQPGVKLGGEAAHLPHLRPLPESQSPSPILLQTAPGVKGHQRVVTLAQ--HISEVI 2143
Db 1436 LHLHQDPLHQGSAGPVHPLVDPL-----TAGPHLARFPYPGTLPLNLLGQPPHEHEIV 1490
Qy 2144 TQDTRHHPOQLSAPLAPLYSFPGASCPLDLRRPPSDLYLPPPDHGAP-----2193
Db 1491 -----RHP-VGTYPRDL---PGA-----IPPMGSAHQLOQAHQAOSA 1525
Qy 2194 -----ARGSPHSEGG 2203
Db 1526 ELQRLAMEQQLHGHHPHMG 1546

RESULT 57
QYR40
ID Q7YR40 PRELIMINARY; PRT; 2171 AA.
AC Q7YR40;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Homologue to Drosophila photoreceptor protein calphoton.
GN KIAA0170.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22709134; PubMed=12799463;
RA Anzai T., Shiina T., Kimura N., Yanagiya K., Kohata S., Shigenari A.,
RA Yamagata T., Kuleki J.K., Naruse T.K., Fujimori Y., Fukuzumi Y.,
RA Yamazaki M., Tashiro H., Iwamoto C., Umenaha Y., Imanishi T.,
RA Meyer A., Ikeo K., Gojobori T., Bahram S., Inoko H.;
RT "Comparative sequencing of human and chimpanzee MHC class I regions
RT unveils insertions/deletions as the major path to genomic
RT divergence."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7708-7713(2003).
DR EMBL; AB100084; BAC78176.1; -.
KW Receptor.
SQ SEQUENCE 2171 AA; 235386 MW; 8DF30C9CDEF7114 CRC64;

Query Match 3.4%; Score 452.5; DB 6; Length 2171;
Best Local Similarity 19.8%; Pred. No. 3.1e-13;
Matches 495; Conservative 290; Mismatches 897; Indels 817; Gaps 116;

Qy 18 RYPHSLSYFV-QIARTHTDVGLEYQHHS--RDYASHLSFGSIITQQRRRPSLLSEFQP 74
Db 58 RMPDCSVALPFFSISKQHAIEILAWDKAPILRDGSLNGTQILRP---PKVLSPGVS 112
Qy 75 GNERQEL-----HLRPESHY---LPELKSMEFIESKRPLELLPDLRPSLLATG 127
Db 113 HRLRQELTLFADLLLCQHLRDLVSLFVSRLGTVEETPRVQGGTQQLR-----LTAED 167
Qy 128 QPAGSEDLTKDRSLATKGLFVSPSPPTDDELE-----LVPPRL-----SKEELIQNWD 177
Db 168 SEEVDFLSRRVWKSRRTSSVIVPESDEGHSPLVGLGLPFPFNLNSDITDVEEQQ 227
Qy 178 RVDREITWVEQISLKKKQKQLEBEAAKPPPEKVPSPPTIESKHSILVQIYDENRKK 237
```

```
Db 228 PATESSAARRGATVEAKQSEAEVVTETIQLEKQDP-----LVKRRDDDTKVK 275
Qy 238 AEAARILEGLGQVELPLYNQ-----SDTROXHENIKINQAMRKKLILYFKERNHARKQ 293
Db 276 RGEN-----GVVPAGVILERSQPPGEDSDT-----301
Qy 294 WKOKFCORYDQMLEALEKKVERIENPRRRAKESKRYEYVEKOFFEIRKORELOER-----349
Db 302 -----DVDDDSRPPGPAEAVHLERAQPPGFINSDTDAEERIPA 340
Qy 350 -----MQSR-----VGQSG-----LSMSARSEHEVSEIILGLSEQENLEKQRL 392
Db 341 TPVVIWMKKEKIFHGVTGPGAPGLAHLQESQAGSDTVEE---GKAPQ-----387
Qy 393 AVIPPMLYDADQRIKFINMNLMADEPMKYKDRQVMNMWSEOEKETFREKFMHOKNFG 452
Db 388 -----NVPEKSQASWINSDTDEEVEVSAALTLAHLK---420
Qy 453 LIASFLEKRTVABCVLYLYLTKKENYKSLVRRSYRRRGKSOQOQOQOQOQOQOQOQPM 512
Db 421 -----ESQPAIMNRDAEDM 435
Qy 513 PRS---SQEKKDEKEKEAEKEEKEPEVENDKEDLLK-----EKTDTSGE 556
Db 436 PORVLLQRSQTTTERDSDTDEVEELPVEN-REAVLKHOTKTRALVRAHSEKQDPPFGD 494
Qy 557 DNDEKEA-VASKGRKTANSQGRKGRITRSMANEANSEAI-----TPQOS-- 601
Db 495 SDDSVEADKSPGIHLERSQASTTVDINTQVEKEVPPGSAIHLKHKQVSEGTNQTDVK 554
Qy 602 -----AELASMELNESSRW-----TEEMETA-----KKGLL-----BHRNWSA 636
Db 555 AVGGPAKLLVVSLEEA--WPLHGDCTDAEEDTSLAASAVADVRSKSLPAEGDAGAEWAA 612
Qy 637 I-----ARMVSK---TVSQCKNFYNYKKRONLDEILQOHLKMEKERNARKKKKAP 687
Db 613 AVLKQEAHEVGAQGGPPVAQEQDL--PISRENLTDLVVDTDTLGESTQFOR-----663
Qy 688 AAASEEAAFPVVEDEMEASGVS-----GNEEMVEEA-----EALHASGNEV 731
Db 664 -----EGAQVPTGREHQHVGKTDSEDTGDSDDLQATQCFLENQGLEAVQSMEDP 718
Qy 732 PR-----GECS-----GPATVNSSDTESIP-SPHTEAAKDT 762
Db 719 TQAFMLTPPQELGSHCSFOTGTGLDBPWEVLATQPFCLRESEDSPTQPTDTHLEA---774
Qy 763 QGQCP--KPAITLGDGPPPGPTPPRTSRAPTEPTAS-EATGAPTPPAPSPSAPP 819
Db 775 --YGFCLSPRAIPGDQHP-----ESPVHTPEMGIQGRGTVDKVMG-----815
Qy 820 PVVPEKEEBETAAPVVEEGEOKPPAAEBELAVDTGKABEPVKSECTEEAEEPKAGKD 879
Db 816 --IPKETAARVGPGRFLERETEKLLP---ERQDVTGEEELTKGQDRQKOLLARDTQ 870
Qy 880 AEAATAEGALKAEKKEGGSGRATTAKSGAPQSDSSATCSADEVDEAGGDKNLLS 939
Db 871 QESDKNGESASPERDRESLKVEIETSEEIQEKQVQKTLPSKAFE-REVERPVANRECD 929
Qy 940 PR-----PSLL-----TPTGDPANASPOKPLDLKQLKORAAATPPIQTVKHHEPPREDA 989
Db 930 PALEEEKVPKILERTQGEPEGSGDQK-----GQASSPIP-----BFGVE-- 972
Qy 990 APTKAPPAPPQNLOPES-DAPOQGSPPRGSRSPAPPADKAEFAAEAAQKLP-GDPP 1047
Db 973 AGDLPGTSAVPTSGSGSGGSGSPVSRHCKGLLNCKMPPAEKASRIRAEKVSRGDQE 1032
Qy 1048 CWTSGLPFPVPRVVIKASHPADPSAFSYPAPPGLPLGLHDTARVPLPPPTISNPPP 1107
Db 1033 SPDACLPPTV-----PEAP-----APPOKP-----LNSQSKHLAPPLLS---P 1069
Qy 1108 LISSAKHPSVLE-BOIGAISQMSVOLHVPVSEHAKAPGVPTWGLPLPMDPKKLAPPSG 1166
```


Db 1070 LLPSTK-PTVRKTRQGS-----QEAPEAPLS-----SELEPH- 1102
Qy 1167 VKQEQLSPRGQAPGESLGVPTAQBASVLGRTALGSLVPGGSIKIGIPSTRVPSDSA1--- 1223
Db 1103 -----PKPIKTRKSRMTTFPATSAPEHPSTSTAQVTPKPTSAQTSR 1149
Qy 1224 TVRSGI---THGTP-ADVLYKGTIT-RIIGEDSPRLDRGRDLSLPGKHVIYEGKKHVL 1278
Db 1150 TNRSSVKTPSEQVPTAPELQPCSTSDQPVTSFTSQVTRGR----- 1191
Qy 1279 SYEGCMSVTCQSKEDGRSSGPPHETAAKRYTDMMEGRVGRASIASIEGLMGRAPPE 1338
Db 1192 -----SRSSVKTP-ETVPTALELQFSTSTDRPVTS-----BPT 1224
Qy 1339 RHPHLLKEQHIRGSIITGIPRSVVEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYK 1398
Db 1225 SHATGRKN-----RSSVKTP-----PVVPTAPELQPSTS 1255
Qy 1399 TQALGPLKPKAHEGLVATVKEAGRSIHEIPRELRHTPEL-----PIAPRPLKEGS 1450
Db 1256 TDQ--PVTSEPTYQ---ATRGKRNSSVKTPPEPVVPTAPELQPSTSDQVTPKPT--- 1306
Qy 1451 ITQGTPLKYDTCASTGSKKHVRSLSGPGTTP-----PVHPLDMADAR 1497
Db 1307 -----SRTTRKTN--MSSVKNPESTVPIAPELPSTSTEQVTP----- 1344
Qy 1498 ALERACYESLSRPGCTASSGG-----SIARGAPVIVPELGPQSPPLYEDHGAPPAGH 1553
Db 1345 -----PFTSATGRKRNSSGKTPETILVTPAKLEPSTS--TDQVTPPTSQATGR 1395
Qy 1554 LPRGSPVTRBPTPLQBGSLSSSKASQDRKLTSTP-----REIAKSPHSTV--- 1600
Db 1396 TNRSSVKTPETVPTAPELQST---STDQAVTPKPTSKTRTNRMTSSVKNPESTVPIA 1452
Qy 1601 PEHHP-----HPISEYHLLRGVGVLYRSHIPLAFDP-TSIPGIPLDAAAAYLPRH 1654
Db 1453 PELPSTSTEQVTP-EPTSRATRG-----RKNRSSGKTPETILVTPAKLEPSTSDQ- 1505
Qy 1655 LAPNPTYPLHYPIRYLIRGVPDPAALENRQTIINDYITSQOMHNTATAMAQRADMLRGLS 1714
Db 1506 VTPEPT-----SQATRGTRNRSVKTPETVTP---TAPELQSTSD---QVPT 1548
Qy 1715 PRESSALNAYAGPRGIIDLQVPHLVLPVPTGTPATAMDRLAYLPTAPQPFSSRHS 1774
Db 1549 PEPTSQV-----TRGTRDRSSVK-----TPETV-----VPTAPELQASATD 1585
Qy 1775 SPLSPGGTHLTKPTTTSSSERDRDRDREREKSLTSTTTVEHAPIWRP--GTE 1832
Db 1586 QPVT-----SEPTSRIT-----RGRKNRSSVKTPETVPTAPELQSTSD 1626
Qy 1833 QSSGSSGSGGGSSSRPASHASHAHQHSPISPRTQDALQORPSVLHNTGMKGIITAVEP 1892
Db 1627 QVTPETSOATRGTRNR-SSVKTPESIVPIAPELQ-----PSTSRNQ-----LVTPEPT 1675
Qy 1893 SKPTVLR---STSTSSVPRPAATPPATHCPGLGTLGCVPTLMEPVLKPEAPRVARP 1948
Db 1676 SRATCRTRNRSVKTPPEVPTAPEPHPT-----TSTDQVPTPKLTSRATR 1722
Qy 1949 ERPRADTGHAFIAPK--PARSGLEPASSFSKSEPRPLVPVPSGHATTARTPAKMLAPHH 2006
Db 1723 KTRNSVK---TPKPEVPAASDLPEFTPTDQSVTPPEAIAQ--GGOSKTLRSTVRAMPVP 1777
Qy 2007 ASPDPAPPASADPHREKTQSKPFSIQEELRSLSLYGHSSYSPGVEPVSPVSPSLTH 2066
Db 1778 TTYPEQSP-----VTTDQPI-----SPB-----PITQPS-- 1802
Qy 2067 DKGLPHLELDKSHLEGELRKPQCPVKLGGAAHPLHRLP-PESQFSSSPLLQTAGP 2125
Db 1803 -----IKRQRAAG-----NPGSLAIPIDHKPCSAFLEPKSQASRN---QRNGA 1842
Qy 2126 VKGHQVVTLAQIHSEVITQDYTRHHPOOLSAPLAPLYSPFGASCPLVDLRRPPSDLYL 2185
Db 1843 VRAESLTAIPAPASPOLLE--TPIHASQIQKVEPAGRSRFTPELQPKASQSRKSLATM 1900

Qy 2186 PPPDHGAPARGSPHSEGGKRSPEPNKTSVLGGEDGIEFVSPPEGWTBPGHRSRAVYP-- 2243
Db 1901 DSP-----PHQKQQRGEVSQKTVIKEEEE-----DTAEKPGKEEDVVTPKP 1943
Qy 2244 -LLVRDGTQTPSRMGSKSPGNTSOPPAFFSKLTESNSA 2281
Db 1944 GKRRDQAEEBPNRIPSRSLRT-----KLQOESTA 1974
RESULT 58
ID Q9V5L9 PRELIMINARY; PRT; 2971 AA.
AC Q9V5L9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Transcriptional activator SRCAP.
GN SRCAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99278407; PubMed=10347196;
RA Johnston H., Kneer J., Chackalaparampil I., Yaciuk P., Chrivia J.;
RT "Identification of a novel SNF2/SWI2 protein family member, SRCAP,
which interacts with CREB-binding protein.";
RL J. Biol. Chem. 274:16370-16376 (1999).
DR EMBL; AF143946; AAD39760.1; -.
DR TRANSFAC; T04151; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008086; F:ATP dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF02178; AT hook; 3.
DR Pfam; PF00271; helicase C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR PRINTS; PR00929; ATHOOK.
DR SMART; SM00384; AT_hook; 3.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICc; 1.
KW ATP-binding; Helicase; Hydrolase.
SQ SEQUENCE 2971 AA; 315639 MW; F9F7EE70304B78A3 CRC64;

Query Match 3.4%; Score 452; DB 4; Length 2971;
Best Local Similarity 19.0%; Pred. No. 4.8e-13;
Matches 587; Conservative 296; Mismatches 1107; Indels 1092; Gaps 130;
Qy 71 EFQGNERSQ-...LHLRPESHSLVPELGKSEMEFI--ESKRPRLELLPD--PLLRSPP 122
Db 91 DFQOEDEEEDDEETIEVEEQEGNDAAQRRTELLRRREGELPLEELRLSPOLLSEGP 150
Qy 123 LLATGQAPAGSBDLTKDRSLATGKLEFPVSPSPPHDTPPELVPFR---LSKEELIQNMDRV 179
Db 151 SSPSTPSSHSDTRDGPGEAGEE-----EPQV---LEIKPPPSAVTQRNKQPHWPD 202
Qy 180 DREITWVEQIISKULK---KKQQQLBEEAAKPEPEKPVSPPIESKHSLSVLQIIV---- 231
Db 203 DEEFTANEDEEEDTIAAEQLEGEVDHAMELSAREGEL-----SMEELQOYAGA 257
Qy 232 -----DENRKKAAAHRIELGLGP-OVELPLYNQPSDTQYHNIKINQAMRKXLI 281
Db 258 YAPGSGSEDEDEVEDANSDDCEPGEVEAEEP---PQEDSSSSQSDSVE----- 304
Qy 282 LYFKRRNHARKQKQKFCQRYDQLMEALEKKKVERIENNPRRRAKESKVREYEQFPEIR 341

Db 305 -----DRSEDEHSEEEETSGSASESESESEDAQSOSA 343
Qy 342 KORE-----LOERMQRVQORSGLSMAARSEHEVSEIIDGLSEONLEKOMRQ 391
Db 344 DEEBEDDFGVEYLLARDEQSEADAGSPPTPGTTLPGPKETIDIAAAAEISLQPKGYT 403
Qy 392 LAVIP-----PMLYDADQORIKFI-----NMGLMADPMKYVKORQVNMW--- 431
Db 404 LATTQVKTPIPLLLRQGLREYCHIGLWLVMTYKKNGLILADEMGLKTIQIISLAHL 463
Qy 432 -----WS-----EOKETFRKFMQHPKNGFLIASFLERKTVAR----- 465
Db 464 ACEKGNWGHLLIIVPTSVMNLWEMELKWCSPF---KILTYGAQXERKLKRGWTKPN 519
Qy 466 -----CVLYYYIT-----KKNENYKSLVRRSRRGKSQOQOQOQOQOQOQOQO 507
Db 520 AFHVCIITSYKLVLDHQAFRRKNRYLLIDEAQNIKNFKSQWOSLLNFNSQORLLLTGT 579
Qy 508 -----QOQPMPSQOEKDEKE-----KEKEAKEEB-----KP- 536
Db 580 PLONSLMELMSLHFMHVFQSHREKFWNSNPLTGMIEGQEVNEGLVRLHKLVRPF 639
Qy 537 -----EVENDEKDLLEKTD-----DTSGEDNDEKEAVASKGRKTANSQGR 578
Db 640 LLRRVKVDVEKMPKYEHVIRCLSKQRCLYDDFMAQTITKETLAT-----GHF 690
Qy 579 KGRIT-----RSMANEAN-----SEBAITQOQSAELASHELNESSRWTEEMETAKG 626
Db 691 MSVINILMQLRVCNHNPLDFPRVTPSTPITPGICFSTASIVLRATDVHPLQRIQDMGRFD 750
Qy 627 LLE-HGRNWSAIFARMVGSKTYSQCKNFYFNKYKRONLDEILQOHLKMKERNARRKKX 685
Db 751 LTGLEGR-----VSRYEADTFPLRHL-----SRVULLE 779
Qy 686 APAASEEAAFPVVEDEMEASGVSGNEEVEAEALHAGNEVPRGEGSPATVNNNS 745
Db 780 VATADP-----PVRKPKVMKN-----RMLQ-----PVPKQEGRTVVVNNP 818
Qy 746 -----SDTESIPSHTE-ANKDTGQNGPKP-----PATLGADGPPGPP-----TPPRRTSRA 792
Db 819 RAPLGFVVRPPPGPPELSAQPT--PCPVQVLPASLWVSASGAPPLIPASRPPGFVLLP 876
Qy 793 PIEPTPASEATCAPT-----PPAP-----PSPGAPPVVPKKEEBETAAPVVE 838
Db 877 PIQPSGSLPQVLPGLVGLSTSRPPTLSLKTTPAPVRLSP-----APPG 926
Qy 839 EGEEQKPPAAEELAVDTGKAEPVKSECTEEAEEGPAKGDAAEABATAGALKAEKKEG 898
Db 927 SSSLLKP-----LTVPFGYTPFP----- 944
Qy 899 GSGRATTAKSGAPQSDSSATCSADEVDEARGDKNRLSPRSLLTPTGD----- 950
Db 945 -----AAATTT-----STTATATTAVPAPTAPAFORLILSDMOARLPSGEVWSIGOLA 994
Qy 951 -----PRANASPOKPLDLK-----OLKQ-----RAAIPPIQVTK-----V 981
Db 995 SLAQRVANAGSKPLTFQIQGNKLTGCAVQRLAVGQRPLOMPPTWNTVGVKIIV 1054
Qy 982 HBPREDAAPTKAPPAPPPQNLOPESDAPQ---OPGSSPR---GKGRSPAP----- 1028
Db 1055 RQAPRDGLTFVPLAPAPPPSSGLPAVLNPRPTLTPGLRPTLTGTARAPMPTTLVRP 1114
Qy 1029 -----PADKEAFAAEAKLQCDPCWTSGLFPFVP-----PREVIKASHAPDP 1072
Db 1115 LKLVHSPSPVESASA-----PGAAFL-TISSPLHVPSSLPGPASPMPDPNSSPLASPV 1168
Qy 1073 SAFSYAPPGHPLPLGLHDTARVPLPRPPTISNPPPLISSAKHPSVLERQIGAISQMSVQ 1132
Db 1169 SSVSVPLSSSLPISVPTTLPAASAPLTIPIASPLTVSASGALL----- 1214
Qy 1133 LHVPYSEHAKAPVGPVTMGLPLDMDPKLAPPSGVKQEQLSRPGQGPES---LGVPTA 1189

Db 1215 -----TSVTPPLAPVVPAPGP-----PSLQPSG-ASPSASALTILGLATA 1253
Qy 1190 Q-----EASVLRTALGSPVGGSIKGIPISTVRPDSAITYRGSIHGH----- 1232
Db 1254 PSUSSQTTGHPHLLAPTSVHVPGLNSTVAPACSPVLVPASALASFPFAPNAPAPAQASL 1313
Qy 1233 -TPADVLYKGTITRIIGEDSPSRDLRGREDSLPKGHVIV-----EGKKGHVLSVEGMSVT 1287
Db 1314 LAPASSAQALATPLAPMAAPQTAILAPSPAPPLAPLVPPLAPSPGAAPVLASSQTPVPM 1373
Qy 1288 QCKEGRS---SSGP---PHETAAPKRTYDMMEGRVRAI--SSASIEGL-MGRAIPERH 1340
Db 1374 APSSTGCTSLASAPVAPPTVPLAPSTQTMPLPAPVPSPLPSPASTQTLALAPALAPT-- 1431
Qy 1341 SPHLKQHQHRSITQGIPIRSYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAYKTQ 1400
Db 1432 -----LGGS-----SPSQTLISLGTGNPQGPFTQTL 1458
Qy 1401 ALGFLKLPKPAHEGLVATVKEAGRSIHIEPREELRHTEPLAPR-PLKESGITQGPLKY 1459
Db 1459 -----LTPA-SSLVPT---PAQTLSLAPGPPPLGPTQTLSLAPPLAPSPVGPAPHT 1508
Qy 1460 DTGASTGSKK---HDVRSIGSPGRTPPVPHPLDVMDADARALERACYEESLKSREGTA 1515
Db 1509 LTLAPASSASLAPASVQTLTSPA-----PVTILGPAQAQTLALAPASTQSPASQA 1561
Qy 1516 SSSGGSITARGAPVIV-----PELGKPRQSPLTVEDHGAPFAGHLPGRGSPVTMR 1563
Db 1562 SSLVNSGAPPLPVTMVSLPVSKDEPDLTLRSGLPPSPSTATSGGPRRRQP--- 1617
Qy 1564 EPTPR-----LOE-----GSLSSKASQDRKLTSTPREAK-- 1594
Db 1618 PPPRSFYDLSLEKKRQRORSELERIFOLSEAHGALAPVYGVTEVLDFCTLPQVASPI 1677
Qy 1595 SPSTVPEH-----HPFISVYELLRGVSGVDLYRSHIPLAFDPTSIPRGIPLD 1644
Db 1678 GPRSPGSHPTFTWYTEAHRAVLFPQOURLDQSEIIFVWP-----PVEAPPP 1729
Qy 1645 AAAAYLPRHLAP-----NPTYPHLYPPYLI-----RGYPD-----TA 1677
Db 1730 SLHACHFPPLWLAQRAAFQOLASELWPRARPLHRIVCNMRTQFPDLRLIQYDCGKLQTL 1789
Qy 1678 ALENRO-----TIINDYITQ---QMHNTATAMAOADMRLG--- 1712
Db 1790 AVLLRQLKABGRVLIIFTOMTRMLDVLEQFLTGHLYLRLDGDSTRVEQOALMERFNAD 1849
Qy 1713 -----LSPRESSALNVAAGPRGIIDLQVPHLPVLVPPPTGTPATAMDR----- 1757
Db 1850 KRIFCFILSTRSGGVNLGTADTVFYDSW-----NPTMDAQADRCHRIQTR 1900
Qy 1758 -----LAYLPTA-----POPFSSRHS 1773
Db 1901 DVHIYRLISERTVEENILKANOKRMLGDMAIEGNGFTTAYFKQOTIRELFDLPLEBPS 1960
Qy 1774 SS-PLSP-----GQPHLTKEPTTSSSEREDR----- 1800
Db 1961 SSVSAPEEEETVASKQTHILEQALCRAEDEDIRAATAQAEQVLAELAEFNENDGPPA 2020
Qy 1801 -----DRERDREREKSIILTSTTTVEH----- 1823
Db 2021 GEGEACRPGAEDEMRAEQEIAALVYEQLTPIERYAMKFLASLEEVSRLEAKQAEQV 2080
Qy 1824 -----APIWR-PGTBQSSGSG--SSGGGSSSRPASHSHAHQH--SPISPRT 1867
Db 2081 EAARKDLQAAKEEVFRLPQEEEGFGADESSCGTGGTHRRSKKAKAPERFGRVSERL 2140
Qy 1868 QDALQORPSVLHNTGMKGIITAVEPSKTVLRSST---SSVPRPAATFPFATHCPLGTT 1924
Db 2141 RGAETQCANHTP-----VISAHQ-----TRSTITTPRCSPARVPRPAPRPPPTAS 2190
Qy 1925 LDGVVPTLMEVLLPKCAPRVARPERPRADTGHAFKAPPARSGLEPASSKSGSEPRPL 1984
Db 2191 APAALPAL---VVPVSAFVPIAPNPITILPVHILPSPPPPSQIPPCSSPA---CTPPPA 2245

Db 1528 PLAGDEESNLPKLPQDIFEEAPVAVTTAAPSXDDGQKQVEVEEKPIEDQKPIEDT 1587
Qy 1573 -----SLSSKASODKLTSTPREIAKSPHSTVPEHHPHIPSPYEHLLRGVSGVDLYRSH 1627
Db 1588 STPTSSNEIPESDRATTIAPSKEEPEPSTGAPTQDEPAEPSTDAPEDESKEPSE 1647
Qy 1628 IPLAFDPTSIPIRGIPLOAAAAYILPRHLPAPNTTYPHYLPYLRIGYPDTAALENRQTIIN 1687
Db 1648 VP-----ITVAGEKIPSS---ITPDEEPTATAPVAKP-----DEDVEKETSTRIPT 1694
Qy 1688 DYITSQQMHNTAT-----AMQADMLRGLSPRESSLALNAAAGPRGIIDLSQVPHLP- 1741
Db 1695 DAPASSEDENSSTDQIPSEVPKPEKPTAPQTPAQTPBEGDI-VGATAAPTITSDVPPVQRLPE 1753
Qy 1742 -VLVP-PTPG-----TPATMDRLAYLTAQPPFSRRSSSPPLSPGCGTHLTKP 1788
Db 1754 EVLAEIPOPSTETGKQODETTAAPSIDRKE--PYVTEIDBEATTVAPISKD-----EKP 1807
Qy 1789 TTTSSSERDRDRDRERREKSIILTTTVEHAPIWRPGRTEQS-SGSSSGSGGGGGS 1847
Db 1808 TEEKPVQKPTGEPSKEEKEKPIEQDVST--EGPV---STEASEAGSTESS----- 1856
Qy 1848 SSRPASHAHQHSPISPTQDALQORPSVLHNTGMKGIITAVPSKPTVLRSTSTSPV 1907
Db 1857 -----BEVKPSTEGEVAEKP-----EDKQP-----SSTAQ 1881
Qy 1908 RPAATFP-ATHCPILGGTLDGVYTLMEPV-----LLPK-----EAPRV--A 1946
Db 1882 APVTIPEISTELP---AQDGKPTSEAPVDSDEDTSAPEDEKIPSVSGVEGEVETTA 1938
Qy 1947 RPERPRADTGHAFKAPPARSGLEPASS-----PSKGSPPRLVPPVPSGHAT 1993
Db 1939 SPQAAEDE-----LKTPAES--FPSSDTKVPETEKPEDETKADETPESVTQVSDVAT 1991
Qy 1994 IARTPA-----KNLAPHASDPDP-----APPASDPHRKTKOSKPSIOELE--- 2037
Db 1992 STSAPVAGGDIKDEQATASPEBEEIKPTIAPAAEIPQ-----SEKEPVDEQEVESGT 2047
Qy 2038 ----LRISGYHGSVSPGVSPVSPSPSLTHDKGL-----PKHLEL----- 2077
Db 2048 KATPESDQPIDEIAPATSGPIDEASTAAPTKEESTTVASAAAPVHDDBIKDVTTTQP 2107
Qy 2078 --DKSHL-----EGLRKPQPGVFKLGGEEAHLPHRLPLPSQP 2114
Db 2108 VADEKEVAAPQDETKTSIDVSTDSTPAQDDEKQDKTEAPV-----APTIVSSP 2155
Qy 2115 -----SSPFLQAPGVKHQV-----TLAQHISEVITQDTRHHHQQLSAPL-- 2159
Db 2156 TADSAADSSTTVEVPSVEIDTKPMDDIMSQTIAHTADGAASTSTEDQD---APVTV 2212
Qy 2160 -----PAPLYS-----FPGASCPLDLRPPSDLYLPPDHGAPA 2194
Db 2213 SPQDAKTPVPAPQDSKTPSSRAPQDAEIPATATPLDNDKIPAT--VAPQDDGQVPA 2270
Qy 2195 RGSHPSECKRSPBNKTSVLGGEDGI-----EPVSPPEGMTEPGHRSAPV 2241
Db 2271 TAAPLDE-----DKIQTTAAPLDEEKIPSTAAPLDDEKIPAPVSPVVDVEPSEKPAV 2324
Qy 2242 YPLIYRGEQTEPS-----RMGSKSPGNTSQPPAPFSKLTENSAMVK 2284
Db 2325 SE--YDGESETPPVHDVETSTDEPTSDAKLKPTTAPATPSESAPATEAEIVPETAPEL 2382
Qy 2285 SK-----KOEINKLANTHNRPEYNIQSP--GTEIFNMPAITGGLMYRSQAOVE 2334
Db 2383 KEVPEKATEQPELEKETPEKATQPELEKETPEKATE---QPEL-----EKETPE 2430
Qy 2335 HASTNMGLEAIIRKALMGKYDQWEESSPLSANAFNPLNASLPAAMPITAADGRSDHLL 2394
Db 2431 KATEQPELEKEVT-----DKATEQPESVDEKITT-----EPVVKPSLSDTEEDBSV 2477
Qy 2395 TSPGGG-----KAKVGRPSRKAQSP--APGLASGDRPPSVSVSHSEGD-----CNRR 2442
Db 2478 ESEBESADKDKNKEETEDTKGHEPEVPVAVSEIQPSEAEVPTTGHPLFPHLASSTT 2537

Qy 2443 TP--LITNRVWED-----RPSSAGSTPPFPYNPLIMLRQAGVMAS-----PPPPGLPA 2486
Db 2538 TTPAVDDRVGEDEENTVKLSSSTTTSTTESPVSPTSAPSTTTVASQQQPPITPPPYG--- 2594
Qy 2487 GSGPLAGPHHA-----WDEE 2501
Db 2595 -----HAPEYDEYDEE 2606
RESULT 61
Q8CF91
ID Q8CF91 PRELIMINARY; PRT; 4969 AA.
AC Q8CF91;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Piccolo.
GN PCLO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22384373; PubMed=12401793;
RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT "Piccolo, a Ca2+ Sensor in Pancreatic beta-Cells.";
RL J. Biol. Chem. 277:50497-50502(2002).
DR EMBL; AB083478; BAC53724.1; .
DR PIR; PT0546; PT0698.
DR PIR; PT0644; PT0644.
DR MGD; MGI:1349390; Pclo.
DR GO; GO:0045202; C:synaptic junction; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR008899; Znf_piccolo.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF05715; Zf_piccolo; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS0106; PDZ; 1.
SQ SEQUENCE 4969 AA; 540677 MW; 2A3F9D372E8EB2D CRC64;
Query Match 3.4%; Score 448.5; DB 11; Length 4969;
Best Local Similarity 19.8%; Pred. No. 1.4e-12;
Matches 397; Conservative 226; Mismatches 684; Indels 695; Gaps 95;
Qy 688 AAASEEAAFPVVDDEMEASGVSGNEEEMVEEAALHASGNVPRGECSPATVNNSSD 747
Db 23 AGGSGSALHFGIPAGMEADLSQLS--EERERQIAAVMSRAQGLPKG--SVPAALAAES-- 75
Qy 748 TESTPSHTBAKDTQNGKPKPATLGADGPPGPPPTPRR-TSRAPTEPTPASEATCAP 806
Db 76 ----PSMHRKQELSSQ-AFQOP-----GKPPDPGRPPQHGKSKRTTDTFRSE---Q 120
Qy 807 TTPPAPPSP-----SAPPVVPK-----E 825
Db 121 KLPGRSESTISLKESKSRDFKEEYKSMMPGPFSDVNPPLSAVSVVKNKFPFDLISDE 180
Qy 826 EKEBETAAAPPVEGBEQK-----PPAABELAVDTGKAEE-----PVKSECTEE 869
Db 181 AVQETTKQKVAQDKQKSEGITKPSLQQPSKLIQKQPGKQEVIPQDIPSKSVSSQ 240
Qy 870 AEE-----GPAKGDAAEAATAEGALKAEKKGGSGRATT-----AKSSGAPQDSDSS 918
Db 241 AEKTKPQAPGTAKPSQSPQATPAQQA-KPVAQQPGPAKATVQQFGPAKSPAQAGTKGS 299

DR GO: 0005215; F: transporter activity; IEA.
DR GO: 0007242; P: intracellular signaling cascade; IEA.
DR GO: 0006810; P: transport; IEA.
DR InterPro: IPR000008; C2.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001565; Synaptotagmin.
DR InterPro: IPR008899; Znf_piccolo.
DR Pfam: PF00168; C2; 2.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF05715; Zf piccolo; 2.
DR PRINTS: PR00360; C2DOMAIN.
DR PRINTS: PR00399; SYNAPTOTAGMN.
DR SMART: SM00239; C2; 2.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00499; C2 DOMAIN 1; 1.
DR PROSITE: PS00004; C2 DOMAIN 2; 2.
DR PROSITE: PS0106; PDZ; 1.
SQ SEQUENCE 5165 AA; 562296 MW; 3BEF9911E7A57B13 CRC64;

Query Match 3.4%; Score 448.5; DB 11; Length 5165;
Best Local Similarity 19.8%; Pred. No. 1.4e-12;
Matches 397; Conservative 226; Mismatches 684; Indels 695; Gaps 95;

QY 688 AAASEEAAFPVVEDEEAEAGVGNEBEVVEAEALHASGNEVPRGECSPATWNSSD 747
DB 23 AGGSGSALHPCI PAGMEADLSQLS-----EEERRQIAAVMSRAQGLPKG--SVPAARAAES-- 75
QY 748 TESIPSPHTEAAKDTGQNGKPPATLGADGPPGPTPPRR--TSRAPIETPPASEATGAP 806
DB 76 ---PSMRKQELSSQ--APQP-----GKPPDGRPPQHLGSKSRITDTFRSE---Q 120
QY 807 TPPAPPSP-----SAPPVVPK-----E 825
DB 121 KLGRSPSTISLKSRSRTDPKEEYKSMMPGFFSDVNPLSAVSVNKNFPPDLISDSE 180
QY 836 EKEETAAAPVEEGEEK-----PPAAEELAVDTKABE-----PVKSECTEE 869
DB 181 AVOEETTKKQVAKQDKSGSEGITKPSLQOQSPKLI PKQOQPGKEVIPQDIPSKSVSSQ 240
QY 870 ABE-----GPAKGKDAEAAATARGALKAKKGGGSGRAT-----AKSGAQDSDSS 918
DB 241 AEKTKPAPGTAQPSQSPQAPQPAQA--KPAQAQPGPAKATVQPGPAQAGTGKS 299
QY 919 AT--CSADEVDEARGGKNLLSPRLTPT--GDPRANASP-----QKPLDLKQLKOR 969
DB 300 PAQPPVTAQPAQAQAGLEKTSLQOQPKSLAQTPQCGKVPQGPAPKSPAQOQCTAKLPQQ 359
QY 970 -----AAAIPIPIQVTKVHEPPREDAAFTKAP-----PAPPPQNLQPSDAPQP 1015
DB 360 PQPQTAQVPGTKTPAQLSGPGKTPAQOQPGTKPSPOQPIPAKPOQPPVATKPOQPO 419
QY 1016 G--SSPRGKSRSPADPAKEAFAAEAKLPGDPPCWTGSLPPVPVPPREVIKASHPADPSA 1074
DB 420 APAKQPQHPHTPAKQPQOQTPAKQP--PQQP---TPAKPOQPOQTPA-KPOQOQTPAK 473
QY 1075 FSYAPGHPPLGLHDTAPVLPPTTSNRPPLISSAKHPSVLERQIGALSQMSVO-L 1133
DB 474 ---POPOQTP-----AKP-QPOQGLGKP-----SAQQPS-----KSIQTVTGRPL 512
QY 1134 HVPYSEHAKAPVPTWGLPMDPKLAPFSGVKQEQLSRGAQGPESLGVPTAQBAS 1193
DB 513 QAPPTSAQAP-----AQGL-----SKTICPLCNTTELLH-----TPEKANFNCTECQ 557
QY 1194 VLRGALGVSGGSTTK-----GIPSTRVSDSAITYRGSIHTGT 1233
DB 558 STVCSLCGFNPPLHTEIKEMCLNCOMQALGELAAIPSPQTPKAASVQPAATASKS 617
QY 1234 RADVLYKGTITRIIGHDSRLDRGDSLRKGVHVIYEGKGHVLSEYEGMSVTCCKED 1293
DB 618 PV-----PSQAAPKE-LP-----SKQD 635
QY 1294 GRSSGPPHETAAPKRTYDMGREGVRATSSASIEGLMGRATPPREHSPHLKQHHIRG 1353

Db 636 -----SPK-----APESKKPPPLVKQPTLHG 656
QY 1354 SITQIPRSVVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTOALGPL-KLKAPE 1412
DB 657 FTATAPQPPV-----AEALPKAPPKP-SAALPEQAKAPVADVEPKQP 700
QY 1413 GLVATVEAGRSIHEIPRELRHTPELPLAPRLKEGSIQTGTPPKYDTGASTGSKGHD 1472
DB 701 KTTETLTDSPSSAAATSKAI-----LSSQVQAQVTTAPPLKTDSAKTS----- 746
QY 1473 VRSLIGSPGRTFPVPHPLDMADARALERACVEESLSKSPGTASSGGSIARGAPVIVPE 1532
DB 747 -----QSPPTGDTITPLDSKAMPASDSKIVSHPGTSEKD-----PVQKKE 791
QY 1533 LGKPROSLTYEDHGAFPAHLPRGSPV-----TWREPTRLQEGSLSSSKASQDKLT 1586
DB 792 EPKAKQTKVTPKDTKP-----VPKSPFTSGTRPTTGQATPQQ-----PPKPEQERRFS 844
QY 1587 STPREIAKSPHSTVPEHHPHIPISPYEHLHGVSDVLYRSHIFLAFDPTSIPRGIPLDAA 1646
DB 845 LNLGGIADAPKS-----QPTTPOETVTKLFGFG----- 873
QY 1647 AAYLPRHLAPNTYPHLYPPYLIIRGYDPAALLENQTIINDYITSQOMHNTATAMAOR 1706
DB 874 -----SQTAAQGPSTGQHPSAKATTAVKKETG-----PAEEN 945
QY 1707 ADMLRGLSPRESSLALNYAAGPRGIIDLSOVPHLPVLVPPT-PGTATAMDRLAYLPTAP 1765
DB 881 SNLIS-----TAG-----QOAPH-----POTGPAAS-----KQAP 906
QY 1766 QPFSSRHSSSPLSPGGP-----THLTKPTTSSSERDRDRDRDREREKSIILSTTT 1820
DB 907 PP-----SQTAAQGPSTGQHPSAKATTAVKKETG-----PAEEN 945
QY 1821 VEHAIPWRPOTBOSSSGSSGGSSSPASHSHAHQSPISPRTDALQOQPSVLHN 1880
DB 946 LEAKPVQAPTAKAEKDKKP-----PGKVSKE-----PTEPE-KAVLAOKPKD--- 989
QY 1881 TGMKGITITAVEPSKPTVLRTSTSTSSPVRPAATPPATHCP-LGTLGDGVYPT-----LME 1934
DB 990 -----TTKPKPACPLCRTELNVGSQDPP--NFWCTECKNOVCNLCGFPNTHLTIQE 1041
QY 1935 PVLLPKAPRVARPERPRADTGHAFKAPARGSGLEPASSFSGK-SEPRPLVPPVSGHA- 1992
DB 1042 WCLNLCQTRAISSQLGDM-----KMPDASSG--PKASVPAPAPAEPPQKPTTAHAK 1093
QY 1993 -----TIARTPAKNLAHPHASPD--PPAPPASADPHREKTSQKPSIQELESLRGVHGS 2046
DB 1094 GKKEETVKAETEKOIPEKETPSIEKTPPAVATD-----OKLE-----E 1132
QY 2047 SYSPGVEVPSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQPGVKLGGEAAHPLH 2106
DB 1133 SEVTKSLVSVLPEKKPS-EBEKALPADKE-----KKPP-----AAEAPPLEEK 1175
QY 2107 RPLPESQ-----PSSPILQTAGVKGHORVV-----TLAHISEVITQDTRHHPQ 2153
DB 1176 KPIBDDQKLPDAPKPSASEGEKRDLLKAHVQIPEEGPIGKVASLACEGEQPDTR--PE 1233
QY 2154 QLSAPLAPLYSPGASCPVL--DLRRPPSDLYLPPDHGA-PARGSPHSGGKRSPEPN 2210
DB 1234 DL-----PGATPQTLPKDRQKESRDVTOFQAGTAKGEGRGSFKDRTEKEDKS 1282
QY 2211 KTSVLGGGEDIEPVSPPEGMTEPHGSRSAV-----YPLLYRDE----- 2250
DB 1283 DTS-----SSQPKS-PQGLSDTGYSDGISGSLGEIPSLIPSEKDLKGLKDKSFSQ 1335
QY 2251 -----QTEPSRMGSKSPGN--TSQPPAPFS 2273
DB 1336 ESSPSSPDLAKLESTVLSILEAQASTLVGEKAEKTKTPQKVSPEQPDQOQKTQTPSETR 1395
QY 2274 KLTESNANVKSKEOEINKLNT-----HNNEPE-----YNISQGTIF 2314
DB 1396 DISIEEIKESQEKVKTSKDSAGQGFSPRKEHENFELVDLSPRASIVSDVEDSSE 1455

```
Qy 2315 NMPAI---TGTLMTYRSQAOVHASTNMGL-EAIRKALMGKYDOWEESPLLSANAFNP 2370
D 1456 NSPVARRKRRTSGSSSEYKQSDSQSGDEDFIRKQII-----EMSADE--- 1502
Qy 2371 LNASASIPAAPIITAADCRSDHTLTSFGGKAKVSGPSSRKAKSPAGGLASGDRPPSV 2430
D 1503 -DASGS-----EDEFIRSOLKEIGGVTEQKRETKGKGSKA---GKHRLTRK 1549
Qy 2431 SSVHSEGDNCRRRTPLTNRWED 2452
D 1550 SSTSFDDAGR-----HSWHD 1566

RESULT 63
ID Q86WG6 PRELIMINARY; PRT; 2948 AA.
AC Q86WG6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Transforming acidic coiled coil 2, long isoform.
GN TACC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22508191; PubMed=12620397;
RA Lauffart B., Gangisetty O., Still I.H.;
RT "Molecular cloning, genomic structure and interactions of the putative
RT breast tumor suppressor TACC2.";
RL Genomics 81:192-201(2003)
DR EMBL; AF528099; AAC62630.1; --
SQ SEQUENCE 2948 AA; 309400 MM; 46711FC0ADCCDC CRC64;

Query Match 3.4%; Score 446; DB 4; Length 2948;
Best Local Similarity 19.9%; Pred. No. 9.3e-13;
Matches 499; Conservative 229; Mismatches 865; Indels 920; Gaps 114;

Qy 678 NARRKKKAPAAAEAAFPVVEDEMAEAGVSGNBEEMVEEABAL-----HAGNEVP 732
D 5 NSTSDNRTLSAQTPRSAQPP-----GNSQNIKKQQDTPGSDPHRDASSIG 51
Qy 733 RGECSGPATVNSSDT--ESIPSPH-TEAAKD--TGONGPK-----PPATLGADGPPPGP 782
D 52 SVGLGGFTAESASLDDCLVSPVTEPRKDPQAGPGEGLLPSPPPSOERHPSSSM 111
Qy 783 P---TPPRRTSRAPTEPTPASEATCAPTPPPAPPSPAPPVVPVPEKEE-----BETAAP 835
D 112 PFAECPPEGCLASPAAPEDGPQTSPPREPAPNAPGDIAPFAERDSSTPYQEIAPV 171
Qy 836 -----PVERGE-----BOKP-----PAAEL-AVDTKABEPVKSECTEAAE 872
D 172 SAGREOPKEGQKSSFSFGIDQSPGMSVPVLRPEPMKAPLCGEGDQPGGFESQEEAA 231
Qy 873 G---PAKGKDA-----EAAEATAEG-----ALKAKEGEG 899
D 232 GGFPPAESRQGVASVQVTPPEAPAAQQTGESAVLEKSPKMAPIPODPAPRASDRERG 291
Qy 900 SGRA-----TTAKSGAPO-----DSSSSATC----- 921
D 292 QGEAPPQYLTDLFLRACHLPRNSGNAPEAEVNAASQESCCQPVGAYLPHAEPLWCLP 351
Qy 922 SADEVDAEGGDKNRL-----LSRPSLLT 946
D 352 SPALVPEAGSGKEALDTIDVQGHPTGMRTCKPNQVVCVAAGQGOEGGLVPSPEPSILT 411
Qy 947 PT----- 948
D 412 PTEBAHPASSIASFPAAQIPIAVEPSSRESVSKAGMPVSADAKEVWDAGLVGLERQ 471
```

```
Qy 949 -----GDPRANASPO-----KPLDLKOLKORAAAIPPIQVTK---VHE----- 983
D 472 VSDLGSGEHPGEGVGPVAPSPQERGEHLNTEQSHVQVGPVPPPLPKESHEVQCAP 531
Qy 984 PPREDAAPTKPA--PPAPPPQNLQPESDAP-----QQPGSSPRGKSRSPAPP--- 1029
D 532 PPPLPKAPSSARPPGPTTGAKVHEDSTSPAVAKESRSRSPGDSPPGKEAPEPPDGGDP 591
Qy 1030 -----ADKEFAAABAKLPGDPPCWTGCLPPVPVPPREVIKASPHAD---PSAFSYAPP- 1080
D 592 GNLOGEDSQAFSSKR-----DP-----EVKDELSKPSDAESRDHSSHSAQPPR 637
Qy 1081 -----GH-----PLPLGLHTARVPLPRPPTISNPPPLISSAKPSVLER 1120
D 638 KGGAGHTDGHPSQTAEADASGLPHKLGEED---PVL-----PPVPDGAGETVPE- 684
Qy 1121 QIGAISOGMSVOLHVPYSEHAKAPVPTMGLPLPMDPKKLA PFSGVYKQEQOLSPRGOAGP 1180
D 685 --GAIWEGSGLQPKCPDTLQSRGLGRMESPLTLESEKSDPPP---TPVAEVPKAEQEG- 738
Qy 1181 PESLGVPQAEBASVLRGTALGSPVGGSIITKIPSTRVPSDSAI--TYRGSITHGTADVL 1238
D 739 -----ESTLEIRKMGSCDGEGLTSPDQPRGPACDASRQEFHAGVPHPPQGENL 787
Qy 1239 YK--GTITRIIGEDSPRLDRGREDSLPKGHVIVYEGKKGHVLSYEGGMSVTCSEKDG-- 1294
D 788 AADLGLTALILDQD-----QQGIPSCFCEGWIRGAASEWPLLSSEKHPQPSQAQPSISF 842
Qy 1295 -----RSSGPPHETAAPKRTYDMMGRVCRALISSASIEGLMGRATPPE---RHSPHLKE 1347
D 843 DVLKEQAOQPPEN---GKETSPSHPGFKDQGDADSSQIH-----VPVEPEQDNNLPTHGQ 893
Qy 1348 QHTRIGSITQIPRSYVE-----AQDDYLRRRAKL-----LKRGTPTPP--- 1386
D 894 EQALGSELQOLPKGTLSDTPTSSPTDMWESSLTERSELSAPTRQKLPALGEKPEGAC 953
Qy 1387 -----PPPSRDLTEAYK-----TQALGPLKLPAPHEGLVATVKEAGRSIHE 1427
D 954 GDQGSRRVSPPAADVLDKDFSLAGNFSRKETCTCGGNKSOQA---LADALEEG-SQHE 1008
Qy 1428 IPRELPH-----TPELPLAPRLKEGSIQTGTPLKYDT----- 1461
D 1009 EACQ--RHPGASEAADGCSPLWGLSKREMASGNTEAPPCQPDVALLDAVPCLPALAPA 1066
Qy 1462 --GASTTGSKKHVRSGLISPGRTFPVHPLDVNVADARALARACVEESLKRSPGT----- 1514
D 1067 SPGVTPQDAPETACDETQEGROQVPAPQOKM-ECEATSDAESPKLLASFPSSAGEGG 1125
Qy 1515 ---ASSGGSIARGAPVIVPELGPQSPPLYEDHGAPFAGHLPRGSPVTWREPTPRLOE 1571
D 1126 EAGAAETGGSAGADPGKQQAPEKPEATLS-----CGLLQTEHCLTSGE-----E 1171
Qy 1572 GSLSSSKAS-----QDRKLSTPRETAKSPHSTV-----PEH 1603
D 1172 ASTSALRESQAEHPMASCQDALLPA--RELGGIPRSTMDFTHQAVPDPKELLLSGPPE 1229
Qy 1604 HPHPISPVEHL---LRGV---SGVDLYRSHIPLAFDPTSI PRGIP---LDAAAAAYLPRH 1654
D 1230 VAAEDTYLHVDSSAQRGAEDSGVKAVSSADPRAPGESPCVGPFPPLALENAASL---K 1285
Qy 1655 LAPNPTYPHLYP-----SPRESSLA-----LNYAAGPRGIIDLSQVPHLPVLVLP 1745
D 1286 LFAGSLAPLQPGAAGGEPAPVAQVASSGSPKARTTEGPVDSMPCLDRMPLAKQKQATGEE 1345
Qy 1677 -----ALENROTIINDYITSQMHNTATAMQADMLRGL--- 1713
D 1346 KAATAPGAKAKASGEGMAGDAAGTEGSMERMGFSPQKQGTGGVDTSSEQTATLTFG 1405
Qy 1714 -----SPRESSLA-----LNYAAGPRGIIDLSQVPHLPVLVLP 1745
D 1406 PDFREHTAKIFEKPVILGALATPGKAGAGRSVAGKDLTRPLGPEKLLD--GPPGVDVTL 1463
Qy 1746 PTPGTPATMDRLAYLPTAPQPFSSSRSSSLSPG--GPTHLT-----K 1787
```

Db 1464 PAP--PARLQVEKKQLAGAEIASHLALQDPASDKLLGPAGLTWERNLPGAGVGKEMAGV 1521
Qy 1788 PTTTSSSER-----ERDRERDRDREREKSLTSTTTVEHAPIWRPQTEQSSG 1836
Db 1522 PPTLREDRPBGGAAMPGLGQAYSLERSQELASGLPSAATQELPVERAAAFV-- 1579
Qy 1837 SSGSGGGGGSSSR-PASHSHAHQSPISPTQDALQORPSVLNHTGMKGIIITAVEPSKP 1895
Db 1580 APHSHGEAVQADRIPSGKHQHETISACDSPHGEDG---PCDFAHTGVPG----- 1625
Qy 1896 TVLSTSTSPVRPAATFPFATHCP-----LGG-----TL 1925
Db 1626 HVPRSTCAPSQREVLTIVPEANSEFWTLDTLGGERRPGVTAGILEMRNALGNQSTPAPPT 1685
Qy 1926 DGUYPTLMEP-----VLJPKEAPRVARPERPADTGHAPE-----LAKPPA 1965
Db 1686 GEVADTTPLEPGKVAGAAAGEAGDITLTAETQACASGDLPEAGTTRTFVSVAGDLVLP-- 1743
Qy 1966 RSGLEPASSPSKSGEPRLVPPVSGHATI-ARTPAKNLAPHASPDPE-PAPPASASDPHR 2023
Db 1744 GSCQDPACSDK-----AGMGEGTAAHGDSPAR---PQAAKEQCPGPERPIAGD--- 1789
Qy 2024 EKTQSPFSIOELRLSLGYHSSYPGVEPVSPSLTHDKPLKLEELDKSHLE 2083
Db 1790 -----GKVCVSSPPPEPDTHD---PK-LQLHAPAEELH 1817
Qy 2084 GELRPKQPGVKL-----GGEAAHLPHLRPLPESOPSSPLLQOT 2122
Db 1818 TDRESPPRGPSMLPSVPKADAPRVMDKVTSDTEGAEGT-----ESSPVADDDIQQ 1868
Qy 2123 APGVKGHRVVTLAQHSITQVTRHPQOLS-----APLP-----APL--- 2163
Db 1869 AAPADLESPTLAASYHCDVVGQVSTDLIAQISPAAHAGLPSAAEHVSPAPAGDR 1928
Qy 2164 --YSPGASCVPDL-RRPPSCLYPPDPHAGARGSPHSGGRKSPENPKNTSVLGGGED 2220
Db 1929 VEASTPSCPDPKDLSSRSDSEAFETPESTTPVKAPE-----APPPPPPEVPEPEV 1981
Qy 2221 GIEPVSPGEGTEGHSRAVYLLYRDEGTEPESRMGSKS---PGNTSQPPAF-FSKLIT 2276
Db 1982 STQP--PPE-----EPG-----CGSETVPDPGPRSDSVEGSPFRPFPSHSFAVF 2034
Qy 2277 ESNAMVSKK-----OEINKLNTNRNEPEYNISQPCTEIFNMPAITGT 2322
Db 2025 DEDKPIASSGTYNLDLFONIELVDFTQLEPRASDAKQEGKVNTRRKSTD--SVPIKST 2082
Qy 2323 GLMYRSQAVQ-----EHASTNMGLEAIIRKALMGKYDOWEESPPISANAFNPLNASAS-- 2376
Db 2083 ---LSRSLSLQASDFDGAASSGNPEAV-----ALAPDAVSTGSSASST 2123
Qy 2377 -----LPAAM-----PITAADGRSDHTLTSPGGGKAKVSGRPSRRKAKS 2416
Db 2124 LKRTKPRPPSLKKKQTKTETPPVKTEQOEDESLVPSGENLASSET-KTESAKTEG 2182
Qy 2417 PAPGL-----ASGDRPPSVSS--VHSEGDGNRRRT-PLTN--RV 2449
Db 2183 PSPALLETPLEPAVGPKAACPLDSAGVVPVPSAGGGRVQNGSPVPPVGRKTLPLITAPEA 2242
Qy 2450 WEDRPS-SAGSTPPFPYPLNLRLOAGV-----MASPPPPGLPAGSGPLA 2492
Db 2243 GEVTPSDSGGQEDSPAKGLSVRLRFDYSEDKSSWDNQENPPPTKKGKPEVA 2295

RESULT 64

Q81SF6

Q81SF6

AC

Q81SF6

DT

01-MAR-2003 (TremBLrel. 23, Created)

DT

01-MAR-2003 (TremBLrel. 23, Last sequence update)

DT

01-OCT-2003 (TremBLrel. 25, Last annotation update)

DE

2Mpa_2 protein.

GN

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2269627; PubMed=12381307;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Mutations in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions.";
RL J. Mol. Biol. 323:533-549 (2002).
DR EMBL; AY130758; AAN61518.1; -.
DR PIR; E89066; E89066.
DR PIR; T33247; T33247.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR01064; Crystallin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001023; Hsp70.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig C2.
DR InterPro; IPR003596; Ig V.
DR InterPro; IPR006025; Pept M Zn BS.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00047; ig; 43.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000089; Hsp70; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 11.
DR SMART; SM00408; IGG2; 37.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; Ty_Kc; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR PROSITE; PS00835; IG_LIKE; 38.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS00142; ZINC PROTEASE; 1.
SQ SEQUENCE 18519 AA; 2051869 MW; 6A8441C5D0BA7729 CRC64;

Query Match 3.4%; Score 445; DB 5; Length 18519;
Best Local Similarity 19.0%; Pred. No. 1.1e-11;
Matches 475; Conservative 334; Mismatches 1017; Indels 672; Gaps 109;

Qy 51 SHLSPGSIQ-----PQRRRPSLLSEFQCNERSQELHLRPESHSHVLPGLKSGMEFIE 104
Db 4934 SEVQQAIIIEQKQVVPVETSAPT-----EPTVE-----LKPVESKETSEV--QQVEIIE 4982

Qy 105 SKRPRLELLPDLRLPS-PLLATQAPAGSEDLTKDRSLTGKLEVPSPSPHTDPELELV 163
Db 4983 QKQVVPVETSAPTVEPTVEKLPAPVESKETSEVQQAIIIEQKQVVPVETSAPTVEPTVEKL 5042

Qy 164 PPRLSKE-----ELIQNMDRVDREIT--MVEQQIISKLK-----KKQQQLSEEA 204
Db 5043 KPVESKETSEVQQAIIIEQKQVVPVETSAPTVEPTVEKLPAPVESKETSEVQQAIIIEQK 5102

Qy 205 AKPEPEKPVSP-----PIESKHSRLVQIIDYENKKAEEAHRILLEGLPQVPLYN 258
Db 5103 VPFVETSAPTVEPTVEKLPVESKETSEVQQAIIIEQKQVVPVETSAPTVEPTVE---KH 5159

Query Match 3.4%; Score 445; DB 5; Length 18534;
Best Local Similarity 19.0%; Pred. No. 1.1e-11;

Matches	475; Conservative	334; Mismatches	1017; Indels	672; Gaps	109
Qy	51	SHLSPGSIQ-----PQRRRP	SHLSPGSIQ-----PQRRRP	SHLSPGSIQ-----PQRRRP	SHLSPGSIQ-----PQRRRP
Db	4934	SEVQQAIIIEQKDVVPVETSAPT	SEVQQAIIIEQKDVVPVETSAPT	SEVQQAIIIEQKDVVPVETSAPT	SEVQQAIIIEQKDVVPVETSAPT
Qy	105	SKRPRLLELPPDLRLPS-PLLAT	SKRPRLLELPPDLRLPS-PLLAT	SKRPRLLELPPDLRLPS-PLLAT	SKRPRLLELPPDLRLPS-PLLAT
Db	4983	QKQVVPVETSAPTVEPTVEKLA	QKQVVPVETSAPTVEPTVEKLA	QKQVVPVETSAPTVEPTVEKLA	QKQVVPVETSAPTVEPTVEKLA
Qy	164	PPRLSKE-----ELIQNMDR	PPRLSKE-----ELIQNMDR	PPRLSKE-----ELIQNMDR	PPRLSKE-----ELIQNMDR
Db	5043	KPVESKETSEVQVVEIEIQKDV	KPVESKETSEVQVVEIEIQKDV	KPVESKETSEVQVVEIEIQKDV	KPVESKETSEVQVVEIEIQKDV
Qy	205	AKPEPEKVPSP-----PIESK	AKPEPEKVPSP-----PIESK	AKPEPEKVPSP-----PIESK	AKPEPEKVPSP-----PIESK
Db	5103	VPVETSAPTVEPTVEKLA	VPVETSAPTVEPTVEKLA	VPVETSAPTVEPTVEKLA	VPVETSAPTVEPTVEKLA
Qy	259	QPSDTRQYHENIKINQAMRK	QPSDTRQYHENIKINQAMRK	QPSDTRQYHENIKINQAMRK	QPSDTRQYHENIKINQAMRK
Db	5160	APVESKETSE-----	APVESKETSE-----	APVESKETSE-----	APVESKETSE-----
Qy	319	N--PRRAKESKVREYKQFPE	N--PRRAKESKVREYKQFPE	N--PRRAKESKVREYKQFPE	N--PRRAKESKVREYKQFPE
Db	5186	TSAPTVEPTVEKLA	TSAPTVEPTVEKLA	TSAPTVEPTVEKLA	TSAPTVEPTVEKLA
Qy	377	DGLSEONLEKQROLAVIPPM	DGLSEONLEKQROLAVIPPM	DGLSEONLEKQROLAVIPPM	DGLSEONLEKQROLAVIPPM
Db	5230	--JTEVE-VELFPSKAEVFG	--JTEVE-VELFPSKAEVFG	--JTEVE-VELFPSKAEVFG	--JTEVE-VELFPSKAEVFG
Qy	434	EQBKETPREKPMQHPKNFG	EQBKETPREKPMQHPKNFG	EQBKETPREKPMQHPKNFG	EQBKETPREKPMQHPKNFG
Db	5284	VESKETSEV-----	VESKETSEV-----	VESKETSEV-----	VESKETSEV-----
Qy	484	RRSYRRRGKSGQOQOQOQO	RRSYRRRGKSGQOQOQOQO	RRSYRRRGKSGQOQOQOQO	RRSYRRRGKSGQOQOQOQO
Db	5324	-----ESKETSEVQQAII	-----ESKETSEVQQAII	-----ESKETSEVQQAII	-----ESKETSEVQQAII
Qy	525	-KKEAKEEKEPEVNDKE-DL	-KKEAKEEKEPEVNDKE-DL	-KKEAKEEKEPEVNDKE-DL	-KKEAKEEKEPEVNDKE-DL
Db	5376	VEQKDVTCESIEIKELITE	VEQKDVTCESIEIKELITE	VEQKDVTCESIEIKELITE	VEQKDVTCESIEIKELITE
Qy	579	KGRITRSMANEANSEAIPTQ	KGRITRSMANEANSEAIPTQ	KGRITRSMANEANSEAIPTQ	KGRITRSMANEANSEAIPTQ
Db	5436	EPITVEKLA	EPITVEKLA	EPITVEKLA	EPITVEKLA
Qy	639	RMVGSKTVCCKNPFYNYKRQ	RMVGSKTVCCKNPFYNYKRQ	RMVGSKTVCCKNPFYNYKRQ	RMVGSKTVCCKNPFYNYKRQ
Db	5482	KSVESKETSEVQO-----	KSVESKETSEVQO-----	KSVESKETSEVQO-----	KSVESKETSEVQO-----
Qy	687	--PAASEEAAP-----	--PAASEEAAP-----	--PAASEEAAP-----	--PAASEEAAP-----
Db	5531	VQPAIVEQKQVVPVETSAPT	VQPAIVEQKQVVPVETSAPT	VQPAIVEQKQVVPVETSAPT	VQPAIVEQKQVVPVETSAPT
Qy	739	PA---TVNNSSDTEIPSHT	PA---TVNNSSDTEIPSHT	PA---TVNNSSDTEIPSHT	PA---TVNNSSDTEIPSHT
Db	5589	PTVEPTVEKLA	PTVEPTVEKLA	PTVEPTVEKLA	PTVEPTVEKLA
Qy	796	PTPASEA-----	PTPASEA-----	PTPASEA-----	PTPASEA-----
Db	5642	SKETSEVEPAIVEQKQVSP	SKETSEVEPAIVEQKQVSP	SKETSEVEPAIVEQKQVSP	SKETSEVEPAIVEQKQVSP
Qy	836	PVEGEEOQPPAAELAVDTG	PVEGEEOQPPAAELAVDTG	PVEGEEOQPPAAELAVDTG	PVEGEEOQPPAAELAVDTG
Db	5702	--ETSAPTVEPTVEKLA-----	--ETSAPTVEPTVEKLA-----	--ETSAPTVEPTVEKLA-----	--ETSAPTVEPTVEKLA-----
Qy	888	EGALKAEKKEGGGS-----	EGALKAEKKEGGGS-----	EGALKAEKKEGGGS-----	EGALKAEKKEGGGS-----
Db	5752	EKLAPVESKETSEVQPAI	EKLAPVESKETSEVQPAI	EKLAPVESKETSEVQPAI	EKLAPVESKETSEVQPAI
Qy	934	KNRLSPRP	KNRLSPRP	KNRLSPRP	KNRLSPRP
Db	5812	QKQV--VPVETSAPTVEP--	QKQV--VPVETSAPTVEP--	QKQV--VPVETSAPTVEP--	QKQV--VPVETSAPTVEP--

Db	78	QAPPPETPSQSQSLLQPOQVLDAPAHKLDVALENGISYTLNSPPRPAISPLTD	137	Db	1094	TTAPVRLNVPVIAPELISWLESFKVTVEKLEYIQKTSKFFQILAMSVINRGLKLTUKSMIP	1153
Qy	836	PVEGEBOQP---PAABEL-----AVDTGKABEVPKSECTEEAEGBPAKAGKDAEAEATAE	888	Qy	1568	-----RLO-----EGSLSSKAS-----	1580
Db	138	PVPDPRTREPSVQPLESELETHAADQPKTNGIAHSNPTNRPNNTDLKGRD--TPQVTSV	195	Db	1154	DESSLELVURVPEERIQTGTSKSLDPLDQLLTTFKATMVSTLEAKLDSQKAECLVLSA	1213
Qy	889	GALAEKKEGSGRATTAKSSGAPQDSDDSATCSADEVDEAEGDKNRLLSRPSLLTPT	948	Qy	1581	-----QDRKLT-----STPRE-----IAKSP-----HST---	1599
Db	196	GA-TAGATVGGAAATTVFVGN--STNSTSHFSGLAWDGT-----IAPFNTQAPPA	243	Db	1214	IDQVLGLADKTIIRGTSESEMRNNAEKVLIIPAIRMVAEWQKQVAATPAAPVHATATP	1273
Qy	949	GDPRANAS-PQKPLDLKQ-----LKQRAAAAPPI-----	976	Qy	1600	--VPEHHPHDISP--VEHLRGVGVDLVRSHTPLAFDPTSIIPRGIPLDAAAAYILPRHL	1655
Db	244	VDGHWANMDFSTWDLLOSQSGHILMSFLOSLSVTTQQOQLOQIPLPLPNTVRPSQVSLA	303	Db	1274	GALPANH-HTMAPATAPSAQRIATPESATORANPGCNPSAVPRTITAAAPAP-	1325
Qy	977	-----QVTKVH-----EPREDAAATKP-----	994	Qy	1656	APNPTVPHLYIRGYPDTAALNRQTI-----INDVITSQOMHNTATAMAQRADM	1709
Db	304	QMQLDOANATHPLHLTAQGGIQPPAAAAATAADASLESFARLEFADGVFQMTYALIIG	363	Db	1326	TPAPLPLPRAAP--ISGHPAPSTVNTNNAAGARPVPAAPGPGVGHSGAIGVTAPV	1382
Qy	995	-----APAPBPPQ-----NLOPES	1009	Qy	1710	LRG-----LSPRESSLALNYAAGPRGIIDLQVPH--LPV-----LVPP	1746
Db	364	RQRAWRLAKKEERRAEQYOLKDEYEAGLTPPAPPSDEDARRFSKSYISEEGNMLGES	423	Db	1383	SRPPVTHGVVTAAPTTPR-SNLAASVPAG---SMAQOHTSGAPAGSGNASISRAPP	1436
Qy	1010	DAPOQSGSPRGKSRSPAPPADK-----EAF-AEAQKLPDPCWTSGLFPVPP	1059	Qy	1747	TPGTATAMDRLAYLPTAPOPFSSRHSSSPLSPGPHLTPTTSSSERERDRDRDR	1806
Db	424	DDEDEEGRP--SDNRAPKKRKNMGVSLPVDGFEAESLSMADQI--VDDGKGVPVN	478	Db	1437	STAAPTA-----PTSVPPSVSTVRPMSYVTPGGLTAPPTASSG-----ASAGY	1481
Qy	1060	REVIKASPHADPDSAFSAVPGHPLP-LGLHDTARPVLPRPPTISNPP-----	1106	Qy	1807	DREREKSLTSTTVBHA-----PIWRPGTEOSSSGSSGSGSGSSSRPASHSHA	1857
Db	479	RQVSHTPGAAAVNLASLRPSPHPTFLGHSIPGNIAAKTKAISREHLKIQFNSQAGVF	538	Db	1482	ARFANASTWPAITSGQAMTSAVQSPSPRP--SSLTGGQVPAIAAASSTRPASGVYN	1540
Qy	1107	-----PLISSA-----KHPSVLREQ-----IGAISQGMS-VOLHVP---	1136	Qy	1858	HOHSPISPRTDALQORPSPVLHNTGMKGIIITAVE-----PSKPTVLR--STSTSSPVRPA	1910
Db	539	EAIPLHKNGFCEDEVHYSHDKVILKSGDRLOVKQVVEFVFIINGVAEGKTGAEEYEPETP	598	Db	1541	PPASSLAPSTHSM--PSAVPTTASGAVSTVSSLAATPLPPSPAPRYGSPNNATPTPA	1597
Qy	1137	-----YSEHAK-----APVGPVTMGLPLPMDPKKLAPFSGYKQBQLSPRQOAG	1179	Qy	1911	ATPPPATHCLPGTLGDVYFTLMEPVLKPKEARVARPERPRADTGHAFILAKPARSGLE	1970
Db	599	ARRYSEGCKEMSFDPESHIDVDRRSTSPEDNEVMSVDPDR-----DSDSGLS----	649	Db	1598	TAIPRA--PL-----PAASSVSAPVTGQP-----SFGA--PASVPTP	1630
Qy	1180	PPESLGVPTAOB---ASVLRTGA-----LGSVPGSITKG	1211	Qy	1971	PASPSKSGSP--RPLVPPVSGHATARTPAKNLAPHASPDPPAPPASADPHREXTQS	2028
Db	650	PDEDILLDPADPHMETVEKEAEEDDDDEARSQSQSVKPRPPEFDDMSVPLMDPPKK	709	Db	1631	PTSASQAGPLSQVCPAPAISSIGATPAASI-----PSSA-----	1668
Qy	1212	IPSTRVSDSAITYRGSITHTGPADVLYKGT-----ITRIIG-----EDSPS	1253	Qy	2029	KPFSIQLELRSLGYHSGSSYSPGCVPEVPSVSSPLTHDKGLPKHLEBOKHLEGELRP	2088
Db	710	RPGRPPKNGIMSKREERLRKQAMELAKNQPPGCEPPVKKXVGRPKHPLPEDAPD	769	Db	1669	-----PAALADVT-----	1676
Qy	1254	RLDGREDSLPKGHVIEYEGKKGHVLSEYEGGMSVTQCKEDGRSS--SGPPHETAAPKRTYD	1312	Qy	2089	KQCPVKLGGENAHPLHLPLPE-----SOPSSSPLLOTAPGVKGHQRVVTLAQHI-	2139
Db	770	RPEKRYKPRKK-----NGEEDGASDAE-----KTIKEKREKPKTPPLELRREDYTEE	818	Db	1677	-YPVPOQASAAAARLP-VTPAPAAHTTAQSVAPQVARPVTQSP-----VQSVADHVT	1726
Qy	1313	MME-----GRVGRAISSASTEGLMGRAPPERHSPHLKQHHIRGSITOGIPRS----	1362	Qy	2140	-SEVITQDYTRHHPOQLSAPLAPLYSPGASCPCVLDLRRPPSDLYLPPFDHGAPARGSP	2198
Db	819	QLOKPNKNGYGLIDVLSAAPDGLTKQI-----YKRIQLKYPFYFNVDTKGSWSSVRHN	874	Db	1727	SSQAST---TAHP--VAQSVPRVNSNPTSAAPV-----AAVGTA	1762
Qy	1363	-----VVEAQDYLREAKLL-----KREGTPPPPPPS--RDLTEAKTQALGPKLKP	1409	Qy	2199	HSEGGKSPBPNKTSVILGGEDGIEPVSPGEMTEPGHS--RSAYVLLYRDGQOTEPSPR	2256
Db	875	LIGNDAFKNEETHLWSRVPGIDIDAGKKRKAPSPDHASSLHNFQCHVAPQPM-----P	928	Db	1763	QVASAPVTQAPHRAHLSVSSQSL-PQSVPHAAQAAAHQTPHSASRPV-----PQSVPOS	1816
Qy	1410	AHEGL-----VATVKEAGRSIHEIPREELRH-----TPELPLAP	1443	Qy	2257	MGSKSPGNTSOPPAFFSKLTESNSAMVSKQKQEKINLKNTHNRNEPEYNISQPTGEIFNM	2316
Db	929	PHPGMYHGHGVQSQVHPGTVGQRQSVYTTGQPGASQH--PQHLQTPQPGVPPQOP--P	984	Db	1817	VPQATPOAVPRP-----STSLTPT-----AQOPV--S	1843
Qy	1444	RPLKGSITOGTFLKYDTCASITGSKKHVDVRSLSIGRGT-----PPV-----	1487	Qy	2317	PAITGTGLMTYRSQAQVQEHASTNMGLEAIRKALMGKYDQWEEESPPIISANAFNPLNASAS	2376
Db	985	REYQAAQTSPPAQ-----AQYGTPTTAAQMGSTPAATYSSPVVSRPPMPTVAAQSGA	1040	Db	1844	PAVSGSGVPA--PSAAQSVAPAPV-----SSTPVAATVAPASTVAA	1883
Qy	1498	-----HPLDWM--ADARALERACYEESLKSRRPGTASSGSGSARGA-----	1526	Qy	2377	LPAAMPITAADGRSDHLYTSPGGGKAKVSGRPSRKAQSPAGLASCGRDPSVSSVHSE	2436
Db	1041	TPHSWARQHSILVSGSPQANGIPRV-----NPPATANTGVVAGGAPPAQQAQPTAN	1093	Db	1884	APTTRVTAAPALAS-AATNP-----APVPSQP-QHQITGOAP--AQOORPPAQAPPA-	1933
Qy	1527	-----PVIVPEL-----QKPROSPLTYEDHGAPFAGHLP-----RGSVPTWREBTP	1567	Qy	2437	GDCNRRPTLTNRVWEDRPSASGSTFPFVYNPLIMRLQAGVMASSPPGLPAGSGPLAGPHH	2496
				Db	1934	-----TPTTISAAPPAPPTLAPPPPPPP-----PPTEDPPPPPPPPPPPPPP--	1979

```
QY 2497 AWDEBPKLL 2506
Db 1980 -----PTPLM 1984

RESULT 67
Q63461
AC Q63461 PRELIMINARY; PRT; 2752 AA.
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Proline-rich protein precursor (Fragment).
GN PRP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Spleen;
RX MEDLINE=89066721; PubMed=3198617;
RA Hemichoto K., Peeters B., Dirckx L., Claessens F., De Clercq N.,
RA Heyns W., Winderickx J., Bannwarth W., Rombauts W.;
RT "A single 12.5-kilobase androgen-regulated mRNA encoding multiple
RT proline-rich polypeptides in the ventral prostate of the rat.";
RL J. Biol. Chem. 263:19159-19165(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Spleen;
RX MEDLINE=92250652; PubMed=1577819;
RA De Clercq N., Hemichoto K., Devos A., Peeters B., Heyns W.,
RA Rombauts W.;
RT "The 4.4-kilodalton proline-rich polypeptides of the rat ventral
RT prostate are the proteolytic products of a 637-kilodalton protein
RT displaying highly repetitive sequences and encoded in a single exon.";
RL J. Biol. Chem. 267:9884-9894(1992).
DR EMBL; M86526; AAA41957.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT NON_TER 2752 2752 PROLINE-RICH PROTEIN.
SQ SEQUENCE 2752 AA; 305057 MW; 55930CDE2818F7F8 CRC64;

Query Match
Best Local Similarity 3.3%; Score 438; DB 11; Length 2752;
Matches 564; Conservative 327; Mismatches 1071; Indels 968; Gaps 138;

QY 21 PHSLSYPVQIARTHTDVGLEYQHSHRDYASHLSPGSIQ--PQRRPSPILLSEFQFGNER 78
Db 226 PHSYLP-----LQSQENAEAPHEFEQVELYQMETQTRNP-----EN 263
QY 79 SOELHLRPESHSHYLPDLKXSEMEFTSKRP-----RLELDPDLLRPSPLLATGPAG--- 131
Db 264 SOQ-----EADYFPQSPDEBPLQKEDPAHYQHHTLP-----TITGKPDVIDL 309
QY 132 ---SEDLTKORSITGKLEVPSPSP----- 153
Db 310 KITSPIKVSESLYEETPTQPGPFVEAKLPSPQQQPAETSPTEGESGIDLESSV 369
QY 154 -PHTDPELELVPPRLSKBELIOMDRVREITMVEQIISKKKQOQLE--EEAAKPPE- 209
Db 370 HPQEDSE-EIGLPTLQEDVDVSHLGPVLED-----ESSLESEQVQLPESSEVSGSEN 424
QY 210 -PEKVPSP---PPISEKSHSLVQIIYDNRKKAHAHRLGEGPQVPLPLYN-QPSDTR 264
Db 425 QPEASVQPTVLPLVEQ-----EILFGAPGSSIET---VIE-----TLPIHEIQTQNE 469
QY 265 QYHENIKINQARKKLLIYFKERNARKQWKQFCQVDQLMLEALE--KKVERI--ENNP 320
Db 470 DYGQQLP-NVTVRPVDVALTVTSEPVKE--TESFLAPQEFVHALEYNDVEFPVNEEP 526
```

```
Db 1504 LELITTPQTFEGELS-QTVQES-----TTQNKEPHKEVAVPVPYQAVTPTESQYQAE 1557
Qy 1225 YRGSIT-----THGTADVLTKGTTIR-----ITGEDSPSR-----LDRGR 1259
Db 1558 YQKSLQPLDLLELTVTSETKEAYHSTISKNLAINQVHIOHPNPAETVQPLDLELTI 1617
Qy 1260 EDLSL---PKGHVYIEGKGHVLSEBGMVSQCKEDGRSSGPP--HETAAPKRTYDMM 1314
Db 1618 SSSLQPTAEGELLYSQE-----TVTIQISEPPKQVWTPVPEYQEVAVPAPVQDOA 1667
Qy 1315 EGRVCRATSSASIE-----GLMGRA-----IPPER----- 1339
Db 1668 KYPLSSIVLSNLSLQDELTLSELLEGEAQLTTPDETVMVLPKDRQGIYPDHDHKKHLNLT 1727
Qy 1340 -----HSPHLKEQHIRGSITOGIPRSVVEAQED-----YLRREAK-----LLKR 1380
Db 1728 EVTNQPFLEHTVQOPTIEERSQIOKTKQITEPGCKKVPLAQESSEVTPWPIIKE 1787
Qy 1381 EGTPEPPPPSRDLTBAYKTOALGPLKJK-----PAHEGLVATVKEA-GRSIIHEIPREE 1432
Db 1788 TAFPTP-----HSMALQSLDEKLTIIHSHSPGWTQOHANLKEKSGHTTGKILLDY 1836
Qy 1433 LRHTPELPLAPRPLKEGSIQTGTPLYDTGAST-----TQSKGHVRSLLIGSPGR 1482
Db 1837 AEPNMEIELKHGHLFLTKTTEATTESNTQMTKSLKQVTAFTQNKKSMLPALVESQDE 1896
Qy 1483 TPPP-----VHPLDVNADARALERACYEESLKSRRPGTASSSGGSIARGAPVIVPELGRPRQ 1538
Db 1897 SQPPNNMSLQPLJD-----OELTSSQP----- 1918
Qy 1539 SPLTYEDHG-APFAGHLPRGSPVTWREPT--PRLQEGSLSSSKASQDKLSTPREIAKS 1595
Db 1919 -----HGWPVHIENTEPKIYHVAEPPTGPFVEPELFFLKTKTSRPVQGTATQMAAS 1971
Qy 1596 PH---STVPEHHPHIPISYEHLLRVSQVDLYRSHIPLAFDPTSIPRIGPIDAAAAAYLP 1652
Db 1972 PREMVSRAPEKN-----EAVLSG-PGEDQDESPPNMSLQSLDQBLTLLSSQPHGWIP 2023
Qy 1653 R-----HLAPNTPYHLYPPYLI-----RGYPD-----TAALENROTIINDYITSQ 1693
Db 2024 HPNTHGKIYHVAEPPTGPFVEPDPLFLKTKTSKPVETWILTRTDKSKEMV-----SQ 2078
Qy 1694 QMHNHTATAMAQRADMLRGLSPRESSL-----ALNYAAGPRGIIDLQSVPHLVLVPPTP 1748
Db 2079 SPKVEEAVLPVHGEQESRSPPNMSLQSLQELTLSSQPHG-----WVPH-----PPNTH 2129
Qy 1749 GTPATAMDRLAYLPTAQPFFSRHSSSPLSPGCP-----THLTKTITTTSSSRE 1797
Db 2130 GK-----IYL-----HYAEP--PTGPFVEPDPLFLKTKTSKPVQGTATRMV 2169
Qy 1798 RDRDRDRDREREKSLTSTTTTVEHAPLWRPGTEQSSGSGSGGGSSRRPASH--- 1854
Db 2170 KSPKEMVSLDPENKEAVP-----PAQEGKGESFSS-----PNMSLQSLDHLF 2213
Qy 1855 --SHAHQHSPISPRTOQALQORPSVLH---NTGMKGIITAVERSKPTVLTRSTSTSPVR 1908
Db 2214 MSSQPHGWLPHPKPTDKI-----YLHVAEPPTG-----PFVEPDPLFLATTK-SKPVQ 2262
Qy 1909 PAATFPFATHCPGLGTLDGVPTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSG 1968
Db 2263 GTTT-----EMAKSPKEMVS-QTPEYK-----EAVLSGP-----G 2291
Qy 1969 LEPASSPSKGSBPRPLVPPVSGHATARTPAKNLAPHASPDPP-----APPA 2016
Db 2292 EQDESPS-----PNTLSKLDQGEVAMSSQPHSGVPHPKTPKIKIYHLSIEPPPG 2342
Qy 2017 SASDPH-----REKTQSKPFSIQEILELSRSLGVHSGSYSGEPGVEPVSPVSPSLTHDKGLP 2071
Db 2343 PFVKPTDLILVKTITKSPAETWTPRRIDKLLKEMVPHSPVEEAV-----FP 2389
Qy 2072 KHLEBLDKSHLEGLERPKQPGVKGGLGGEAAHLPHLRPLPESQPSSSLQTPAGVKGHOR 2131
```

```
Db 2390 AHGEGQDES-----GSPNMP-LQPLDQELTLSSQ----- 2418
Qy 2132 VVTLAQHISEVITODYTRHHHPQQLSAPLAPLAFYFPGASCPLDLRLRRPPSPLYLPPPD-- 2189
Db 2419 -----PHGWVPHPN-----TPGKIY-----LHYAEPPTGPFVEPDPLF 2452
Qy 2190 -----HGAPARGSPH-----SEGKRSPEFNKTSVLGGEDGIEPVSPPEGMTEPGHSR 2238
Db 2453 FLRTTKSKPVQSGPKETAKSPKEMVSTPYKEAVLSGPGEDQDESPPN----- 2503
Qy 2239 SAVYPLLYRDEQETEPSRMGS-----KSPG-----NTSOPPAFFSKLTE--SNSAMVKS 2285
Db 2504 ---MSLSKLDQEVMTSSQPHSGVDPHPKTPKIKIYHLSIEPPPGFVFKVTRDILILVKTTKS 2560
Qy 2286 KKQES-----INKLNTNHRNEPEYNIQPGTEIF-----NMPAITGTGLMT 2326
Db 2561 KPAETWTPRRIDKLLKEMVPHSPYE-----EAVPPAHGEGQDESPPNPLQPLDQELT 2615
Qy 2327 YRSQA-----VOEHASTNNGLEAIRKALMGKYQDWEESPPLSANAFNPLINASASUPAMPI 2383
Db 2616 LSSQPHGWVPHPN-----PGKIYHVAEPPTGPFVEPDPLFLRTPKSKPV 2663
Qy 2384 TAADGRSDHILTSPPGGGKAKVSGRPSRRKAKSPAPGLASCD--RPSPVS 2431
Db 2664 ---QGTPTQMAKSP-----EEMVSLSPKNKETVFFPAQKGQDESPPNIS 2706

RESULT 68
O57580 PRELIMINARY; PRT; 1151 AA.
ID O57580;
AC O57580;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE High molecular mass nuclear antigen (Fragment).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98034404; PubMed=9365273;
RA Shimada K., Harata M., Mizuno S.;
RT "A nuclear matrix-associated high molecular mass nuclear antigen,
RT HMNA, of chicken and marked decrease of its immunoreactivity during
RT the progression of S phase."
RL J. Cell Sci. 110:3031-3041(1997).
DR EMBL; D88440; BAA24137.1; -.
DR PIR; T18535; T18535.
FT NON_TER 1
FT NON_TER 1151
SQ SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;

Query Match 3.3%; Score 436; DB 13; Length 1151;
Best Local Similarity 19.7%; Pred. No. 8.6e-13;
Matches 346; Conservative 154; Mismatches 548; Indels 706; Gaps 77;

Qy 752 PSPTHEAKATGQNGKPP---ATLGADGPPG-PPTPBR-TSRAPLEPTPASEATGAP 806
Db 31 PRPKVPIAELHPAOPPPKPVPIGGAPPPGTEPTTTPSKPTDGADAAAPKASAELT--- 87
Qy 807 TPPPAPSPSAPPVVPVPEKEEETAAAPVVEEGEQKPAAEELA----- 852
Db 88 SPPPASFP-PDGKAPSGAGEAE-AGTPPSPQFAGTTPPSQGAAGAPKGDGTAQPSGT 145
Qy 853 ---VDTKAEPEVKSECTEAEAGPAKGAOAAEATAE-----GALKAEKKGSGSRAT 904
Db 146 KSGADGKPAQDVKPAATTAATEARPAASPTVPKATAEATAVTAASQSPKAAATDAAV 205
Qy 905 TAKSSGAPQSDSATCSADEVDEAGGDKNRLSPPSLTPTGDPANA-----S 956
Db 206 TAASQSPKAPATVEVKPAAAVAKEK-----AVTAAAAAPKATAEKAPAPVTS 253
```


RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleby J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celnikier S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banson J., Beeson K.Y., Buesam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleby J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradercky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003558; AAN12045.1;
DR FlyBase; FBgn0052377; CG32377.
SQ SEQUENCE 9196 AA; 1008711 MW; E253273B32DBBE88 CRC64;
Query Match 3.3%; Score 434.5; DB 5; Length 9196;
Best Local Similarity 19.0%; Pred. No. 1.4e-11;
Matches 537; Conservative 386; Mismatches 1083; Indels 819; Gaps 130;
Qy 55 FGSIIQORRPSLLSEFQPNRSQELHLPESHSLP-----ELCKSEMEF 102
Db 3562 KPFLDKPISHPTSLVTSVTGSGKSS---LHPEKPKSVKTEKVLAKPDDISKS---V 3615
Qy 103 IESKEP-----RLELLDPLLRP---SPLLATGQAGSEDLTKDRSLTGKLEP 147
Db 3616 VETDKSPKEYSDDETEDEIE-IPKLPDKPISHPTSLVGTGTFG-----VDKS-----P 3663
Qy 148 VSPSPPTDPELELVPPRLSKEELIQNWDVREITVVEQIQIKLKKQQLBEEAAKP 207
Db 3664 LNPEKPN-----PEKDKVLAKPDDISKSVKTKPIPK-EVSDDETDDEIEIP 3714
Qy 208 PEPEKPVGPP-----PI---ESKHSLSVQIYIDENRKKAFAAHRILEGLGPQ 251
Db 4637 D-----ENDDEIDFPKLPDKPISHPTSL--VTGVTGCGDKSSSLHP--EEXPKSPKXDEK 4687

Db 3715 KPLDKPISQPTSLVTGATFGDGGSPHLPEBKPS-----PEKDKVLAKPDGSSKS 3766
Qy 252 V---ELPLYNQPSDTRQYHENIKINQAMRKGL-----ILYFKERNHARKQWKQKFC 299
Db 3767 VVETDKSPKEYSD-STEDEIEIPKLPDKPISHPTSLVGTGTFGMDKSPHLPEBKNSP 3825
Qy 300 QRYDQMLEALEKKVERIENPRRRKESK--VREYYEKQFPFIEIKQRELOERMOSRVQR 357
Db 3826 EKDD-----EKVLAKPDDSSKVVVTKDKPIKEYSD--DETEDEIEIPKLPDKPI 3876
Qy 358 GSGLSMGAARSEHVSIIIDGLSEQNLEKQMLAVIPMLYDADQORIKFINNGLMA 417
Db 3877 TS-LVTSVTGSGKSS-----LHPEKPKSPKDKVLKPKDDSSKSVVK----- 3921
Qy 418 DPMKVYKDRQVMNMWMSQEKETFE-----KFMQHPKN-----FGLIASFL---BRKT 462
Db 3922 -----TDKPIKEYSDDETEDEIEIPKLPDKPISHPTSLVGTGTFGVDKSPMLNPEKPN 3975
Qy 463 VAECVLYYYLTKNENYKSLVRSYRRRKSGSQOQQOQQOQQOQQOQQOQQOQQOQQO 520
Db 3976 SPEKDKVLAKPDDISKSVVK-----TDKPIKEYSDDETD 4012
Qy 521 DE-----KEKEKAKEKEE-----KPE-----V 538
Db 4013 DEIBIPKLPDKPIQPTSLVTGATFGDGGSPHLPEBKPSPEKDKVLAKPDGSSKSV 4072
Qy 539 ENDK-----EDLLKEKTDGSDNDEKEAVASKGRKTANSQGRKRGRITRSMANEAN 591
Db 4073 ETDKSPKEYSDDETDKIDFPKPKPFKEATPS-----VT 4109
Qy 592 SEEAITPQOQAEALASMLNENESSRWTEEMETAKKGL-----LEHGRNWSAIA 638
Db 4110 PVSTIPDKALDFVSKAECSITTIQNIKLASISITTKDKIVQPIDLSNLKD-DKFP 4168
Qy 639 RMVSGKTVSQCKNFYFNKQRQNLDEILOQHKLK-----EKERNARKKKKAPAAA 690
Db 4169 TSVGDKAKSPRESVKPNLKEYSKDEKPSHPVSLVTSVMGSGDKSPHLPEBKPSPEKK 4228
Qy 691 SEEAAPP-----VVEDEMEASGVSGNR--EEMVVEAEALH-----ASGN 729
Db 4229 DEKVLKPDSSKSVVETDKSPKEYSDDETDDEIDIPKALDKPISHVSLVTSVMGSGD 4288
Qy 730 E-----VPRGECGPATVNNSSDTEIPSPHTEAAKDTQNGKPK- 769
Db 4289 KSPHLPEBKPSPEKDKVLKPDSSKSVVETD---KPIKEYSDDETDDEIEIPKLP 4345
Qy 770 -----PATLGAD-----GPPGPTPTPRRTSRAPTEPTPASEATGAPTPPAPP 813
Db 4346 DKPISHPTSLVGTGTFGVDKSPHLPEBKPSPEKDKVLAKPDDSSKSVVTKDKP--- 4401
Qy 814 SPSAPPVVPKKEKEEETAAAPVVEGEEOKP---PAAEELAVDTGKAEPEVKSEC-TEE 869
Db 4402 -----IPKEYSDDEND-----DEIDFPKLPDKPISHVSLVTSVMGSGDKSPHLPEE 4448
Qy 870 AEGFAKGDAAEAAEATAEGALKAEKKEGSGRATTAKSSGAPQD--SDSSATCSADEYDE 928
Db 4449 KPKSPEK-KDE-----KVLKPDSSKSVVETDKPIKEYSDDE--NDDEIDF 4493
Qy 929 AEGGDKNRLISPRPSLLTPT---GDPRANASPKPLDLKQLKQRAAIPPIQVTKVHEPP 985
Db 4494 PKALDKP--ISHTTSLVTSVTGSGDKSSSLHPPEKPKSPKDKVL-----PK 4539
Qy 986 REDAAPTKPAPPAPPPQNLQPSDAP-----QQFGSSPR-----GKSRSPAPPA 1030
Db 4540 PDDSSKSVVETDKSPKEYSDDETDDEIEIPKLPDKPISHPTSLVGTGTFGVDKSPHLPE 4599
Qy 1031 DKEAFAAEAKLPGDPCCWTSGLPFPVPPREVIKASPHAPDPSAFSAPPCHPLPLGLHD 1090
Db 4600 EK-----PKSPEKDKVLAKPDDSSKSVVTKDKPIKEYSD 4636
Qy 1091 TARVLPRPPTTINPPPLISSAKHPSVLERQIGAISQMSVOLHVPVSEHAKAPVGPVTM 1150
Db 4637 D-----ENDDEIDFPKLPDKPISHPTSL--VTGVTGCGDKSSSLHP--EEXPKSPKXDEK 4687

Db	1801	DTPP--SKDDQSEBEKPKPKKGPYDDDDDDFFPAAPKPAKGSKAEIADRENEMVRRIA	1855
Qy	2288	QELNKKLNTNRNEPEYNIQSOPCTEIFNNPAITGTGLMTYRSQAVQEHASTNMG-----	2341
Db	1859	EEAKRAAEAK-----AAKKGWGTFSWFAKKEAAAAADANAGSSPGK	1900
Qy	2342	LEAIIRKALMGKYDQWESPPPLSANAFNPLNASLPAAMPITAAADGRSDHLTSP----	2397
Db	1901	PIRAKLGEANSFYDP-EQKRWINKNA-SPEDQAAKKSTPPPKGGI PRSSASSPAPMG	1958
Qy	2398	---GGGKAKVSGRPSRKAQSPAGI-----AS-----GDRPPSVSVHSEGDGCRRTPLT	2446
Db	1959	MGVGGSGAPNTPGASAPPTGPPRPAAALMPSASESNVSGPPSAGPLSPSGSN-----	2012
Qy	2447	NRWEDRPSSAGSTPPPYNPLIMRLQAGVMASPPPPGLPAGSGPLAGPHHAWD-----	2499
Db	2013	-----GPSAGLLSGMGGAAMORPASTSTSGPPAGT---SKPLSATSSIDLLGAAPV	2063
Qy	2500	---EEPKPLLCSQY 2510	
Db	2064	RKRGEAKKPKRAARY 2078	
RESULT 73			
Q9BX49 PRELIMINARY; PRT; 1404 AA.			
ID	Q9BX49		
AC	Q9BX49;		
DT	01-JUN-2001 (Tr-EMBLrel. 17, Created)		
DT	01-JUN-2001 (Tr-EMBLrel. 17, Last sequence update)		
DT	01-OCT-2003 (Tr-EMBLrel. 25, Last annotation update)		
DE	BGI74L6.2 (NSF: megakaryocyte stimulating factor).		
GN	BGI74L6.2		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID	9606;		
ORIGIN	[1]		
ORIGIN	SEQUENCE FROM N.A.		
RP	Wray P.;		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; ALI13553; CAC36090.1; ..		
DR	InterPro; IPR000585; Hemopexin.		
DR	InterPro; IPR001212; Somatomedin_B.		
DR	Pfam; PF00045; hemopexin; 2.		
DR	Pfam; PF01033; Somatomedin_B; 2.		
DR	PRINTS; PR00022; SOMATOMEDINB.		
DR	SMART; SM00120; HX; 2.		
DR	SMART; SM00201; SO; 2.		
DR	PROSITE; PS00024; HEMOPEXIN; 1.		
DR	PROSITE; PS00524; SOMATOMEDIN B; 2.		
SQ	SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDE5 CRC64;		
Query Match 3.2%; Score 429; DB 4; Length 1404;			
Best Local Similarity 20.8%; Pred. No. 2.4e-12;			
Matches 266; Conservative 136; Mismatches 516; Indels 360; Gaps 54;			
Qy	476	NYNYKSLVRSYRRRGKSKQ-----QQQQQQQQQQQQQQQPMPSRSEKDEKE	524
Db	126	SQTIKTTTRKSPKPPNNKTKKVISEETEEHVSVENQESSSSSSSSSTIRIKRS	185
Qy	525	KEKEAEKE-EKPEVENDKEDLLKEK-----TDDT-SGENDKEAVASKGRKTANSQ	575
Db	186	KNSAANRELQKLVKVDKNKRTKKPTKPPVWDEAGSLDNGDFKVTPTDSTTQHNNK	245
Qy	576	GRKGRITRSMANEANSEBAITP-QQSABLASWELNESSRWTEEMETAKGILLEGRNW	634
Db	246	VSTSPKIT--TAKPINPRSLPNSDTSKETSLTVNKTETVETKTTTNTKQSTDGKEK	303
Qy	635	SATARMVGSKTSYQCNFYFNKYKQRQLNDEILQQHKLKWE---KERNARRKKKKAPAAAS	691
Db	304	TTSAKETQSIETKSADL-----APTSKVLAKPTPKAETTTKGPALTTPKEPTTTPK	356
Qy	692	EEAFAFPVVVDEMEBASVSGNEEMVEBAEALHASGNVPRGCGSPATVNNSDTESI	751


```
QY 1383 TPTPPPPSRDLTEAYKTOALGEL-----KLKPAHEGLVATVKEAGRSIHEIPREELRHTP 1437
Db 1801 AWPSSKKKTVAAKNSNGGIARSPKLLKLAQAQKVKRLEYSDDDISDLEDD 1860
QY 1438 E-----LPLAPPLKEGSIQTGTLKYDTGASTTG-----SKKHVRSI----- 1476
Db 1861 DDDDEDEGVPLSGK--GSGQKPSQPLKPTASSTLTGKGKGKAKKKQVSEEDDGA 1918
QY 1477 -----IGSPQRTPPVHPHPLDVMDAR-----ALERACYE-- 1506
Db 1919 ASDRTRTRGRYAYIEDDDSSGGIKFGVHRPTTPEERQKFIQOEELKMLAEKNA 1978
QY 1507 -----SLKSPGTASSGG-----SIARGAPVI----- 1529
Db 1979 EGAKIAATPRLTLKSGVTASEKRTPGKAAGDSLSTVPLSVIROAKVLIDYLQKGET 2038
QY 1530 VPELGKPSQPLTYEDHCAPAGHLPRGSPVTMTREPTPRLQEGSLSSSKASQDKLTSTP 1589
Db 2039 IGDLDVDESEL--DD-----AELPDDLPEMDAIARMEVEEERQFSAEVAARELPQAE 2090
QY 1590 REIAKSPHSTVPEHHPHPISPVEHLLRGVGV-DLYRSHIPLAFDPTSIPIRGIPIDAAAA 1648
Db 2091 EVLRTSPSKQTSRVMPESPASQTSGLQEPHKEKLPK---PTMHP---PL---LR 2141
QY 1649 YLPRHLAPNPTYPHLYPYLIRGYDPTAALENRQTIINDVTSQOMHNTATAMAORAD 1708
Db 2142 HQFPISAGFSPASLVPPHAAQGMHPLQRHLSQT-----VPPQAMHLLQNALS----- 2192
QY 1709 MRLGLSPRESSIAL-NYAAGRGIIDLSQVHLPL-VLVPPTPGTATAMDRLAYLPTAPO 1766
Db 2193 -----APLGQPLGCGNYGSGP-----NSAQHLPLVMSMPAAAAAHLMQSAVASATAR 2241
QY 1767 PFSSRHSSPLSPGPGTHLTPTTTSSEERDRDRDREREKSIILTSTT----- 1820
Db 2242 PVETA-SGNPASDPKPRGRKKVTP-----RDQLOKQTAATAVTAATSTTTPGSAPS 2293
QY 1821 --VEHAIWRPCTEQSSGSSGSGGGSSRRPASH-----SHAHQHSPI- 1864
Db 2294 EKVAQPLFKPHED-----AAPSAPASQASVITRMPSELLPPAHRNHGPPSG 2340
QY 1865 --PRQDALQRPSPVLHNTGKGIITAVEPSKPTVLRSTSTSSVPRPAATPPPATHCPLG 1922
Db 2341 LYPSSADLAR-----FYGVANQOPI-PAVPGSRS-----PSTSGPPRH----- 2379
QY 1923 GTLDGVYPTLMEPVLPLKEAPRVARPERPRADTGHAFKAPARSGLPEASSPSKGSBPR 1982
Db 2380 -----LRLQMPGLPP-----PHASLRTYGPPLPLRGSGPPTTPTTNSR 2423
QY 1983 PLVPPVSGHATTARTPAKNLAPHASPDPP-----APPASADPHREKTSKPSFI 2033
Db 2424 PAY--LHG-----AEHGGSPGPMGGVFSFSGPPPARHATPHLNPYRAPPI-- 2467
QY 2034 QEELRSIYGVSGSVSP- GVEPVSPVSPSLTHDKGLPKHLELDKSHLEGELRPKOPG 2092
Db 2468 -----YGNPNYSPLGAGPT-----GSMR---PG 2489
QY 2093 PVK-LGGEAAHL--HURPLPESOPS-----SSPLLOTAPGVKGHQRVVVTLAQHISEVIT 2144
Db 2490 ADVYAGRGVSPGYYPPLPLSTPSAHAATSSVIVSAP-----HTLTPTNHSVPLT 2543
QY 2145 QDYT--RHHPOQLSAPLAPLYSPFGASCPVLDRRPPESDLYLPPPDHGAPARSPHSEG 2202
Db 2544 HGKTPPQQTPTQSSGPPA-----AAPPPTTTSSTSHKPLASVITSK- 2587
QY 2203 KRSPEPNKTSVLGGEGDGIIEPVPPEGMTPEGHSRAVYLLYRDGQTFPSRMGSKSP 2262
Db 2588 -----KLTTLEAYPIRKSPIAVADVSPAEPTRSAPIAEDSGSAHDTTRAPSSAT 2639
QY 2263 GNTSQPPAFFSKLTESNAMYKSKQ 2288
Db 2640 G-----TAVVGESGLLKSQRE 2656
```

```
RESULT 76
O75046
ID O75046 PRELIMINARY; PRT; 1268 AA.
AC O75046;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA0458.
GN KIAA0458.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98116662; PubMed=9455484;
RA Seki N., Ohira M., Nagase T., Ishikawa K., Miyajima N., Nakajima D.,
RA Nomura N., Ohara O.;
RT "Characterization of cDNA clones in size-fractionated cDNA libraries
RT from human brain.";
RL DNA Res. 4:345-349(1997).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AB007927; BAA32303.1; -.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008267; F:poly-glutamine tract binding; TAS.
DR GO; GO:0006607; P:NLS-bearing substrate-nucleus import; TAS.
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR000949; ELM2.
DR InterPro; IPR001005; MYB DNA binding.
DR InterPro; IPR000679; Znf GATA.
DR Pfam; PF03154; Atrophin-1; 1.
DR Pfam; PF01448; ELM2; 1.
DR Pfam; PF00320; GATA; 1.
DR Pfam; PF00249; myb DNA-binding; 1.
DR SMART; SM00717; SANT; 1.
DR SMART; SM00401; Znf GATA; 1.
KW Hypothetical protein; DNA-binding; Nuclear protein.
SQ SEQUENCE 1268 AA; 138624 MW; 5B646E2239034554 CRC64;

Query Match 3.2%; Score 424.5; DB 4; Length 1268;
Best Local Similarity 21.3%; Pred. No. 3.5e-12;
Matches 312; Conservative 164; Mismatches 544; Indels 445; Gaps 69;

QY 363 MSARSEHEVSEIIDGLSEQENL-----EKMQQLAVIPPMLYDADQOQRIKFINMGLM 416
Db 63 LRAARSMAAFAGMCDGGSTEDGCVAAASRDDTTLNALNTLHESGYDAGKA-----LQLLV 116
QY 417 ADPMKVYKDRQVMNWNWSEQKETFREKFMQHPKNGFLI-ASFLEKRTVAECVLYLYLTKK 475
Db 117 KKPV----PKLIKECWTEDEVKRFVKGLRQYKGNFFIRKELLPNKETGELITFYIYWK 172
QY 476 N-ENYKSLVRSYRRRKSKSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 527
Db 173 TPEAASSRAHRRHRRQAVFRRIKTRTASTPVNTSPSPSSBFLDLSSASEDDFDSSESEQ 232
QY 528 EA-----BKEEEKPEVENDKE-----DLK 547
Db 233 ELKGYACRHCPTTTSKWHGGRNILLCTOCRIHFYKYGELPPPIKVPDPPFMPKPVK 292
QY 548 EKTDDTDEGNDKEAASKRKTANSQGRK-----GRITRSMANSEAEITPOQS 601
Db 293 EEDDGLSGKHS--MRTRSRGSMSTLRSRKKQKQKASPDGR-----TSINEDIRSSGRNS 345
QY 602 AELASMEINSSRWTEEMETAKGLLEHGRNWSAIAARMVGSKTVSQCKNFYVNYKKQN 661
Db 346 PSAASTSNDS-----KAETVKK-----AKKVEEASSPLKS-----NKR-- 381
QY 662 LDEILQOHLKWEKERNARRKKKAPAAASEAFAFPVVEDEEEMEASVSGNEEMVEEA 721
Db 382 -----QREKVASDTEADRTSSK---TKTOEIRSPNSPSSEGESS-----DS 422
```


Qy	920	T-CSADEVDEAGDGKNRLLS	PRPS----	LTPTGDP	PRANAS	PQKPLDLKQLKQRAAA	IIPP	975							
Dy	921	QCPCL	-----	ENQGLEAVQ	SMEDEPTQAFMLTP	-----	PQF-----	732							
Dy	699	QCPCL	-----	ENQGLEAVQ	SMEDEPTQAFMLTP	-----	PQF-----	732							
Qy	976	-----	IQVTKVHEPP	REDAA	TKPAPPAPP	QNLQPED	APQOQSGSPGRKSRSPAPPAD	1031							
Dy	733	SHCS	FQTTGLDEPWEVLA	-----	TOPFLURE	SEDS	ETQPDFD	771							
Qy	1032	KEAPAA	EAKQLPGDPCWT	SGLPFP	PREVIKASPHAP	OPPSAF	YPAGHPPLPLGLHDT	1091							
Dy	772	LEAYG	-----	PCL	-----	AI	PGQHPS-----	PVHTPEMGLQCR	805						
Qy	1092	ARPVL	PRPPTISNPP	PLISSAKHPS	VLEROIGALISQ	MSVOLHV	VPYSEHAKAPVGPVTMG	1151							
Dy	806	GR-----	-----	QTVDK	VMGIP	-----	KETAERVGP	ERG	829						
Qy	1152	LPLPMD	PKKLAP-----	FSGV-----	KQOLSPR-----	QAGPPE-----	-----	1182							
Dy	830	PLERET	EKLLPERQTDV	TGEBEELTKGQDRE	QKLLARDTQ	QESDKNGES	ASPERDRE	888							
Qy	1183	--	SLGVPTAQEAS	VLRG	TALGSPVGS	SITKGI	PTSRVPSAIS	TYRGSITHGTPADLVYK	1240						
Dy	889	SLKVEI	ETSEIQ-----	-----	EKQVQ	QTLPSKA-----	-----	FE	916						
Qy	1241	GTITRI	IG-----	EDSPRLDR	GBDSLPK	GHVIEG	KGHVLSYEGM-----	SVTQCKSED	1293						
Dy	917	REVERP	ANRECDPAEL	-----	EKV	PKVILERTQ	RGEP-----	EGGSQDKQ	QASSPTPEP	969					
Qy	1294	GRSSG	PPHETAAPK	RTYDM	MEGRVGR	AISSASIE	GLMGR	AI	PPERSPHLKEQH	HIRG	1353				
Dy	970	GVGAGD	LPGT	SAP-----	-----	VPSG	SGGRGSPVSPRRHQ-----	-----	KG	1005					
Qy	1354	SITQGI	-----	RSYVEAQ	EDYLRE	AKLLKREG	-----	TPP-----	PPPPSRDLTEAYK	KTQ	1400				
Dy	1006	LLNCKM	PPAEKAS	IRAEK	VSURGD-----	QESD	PCLPTVPEAP	APPQKPLNSQ	SQKH	1060					
Qy	1401	ALGPLK	LKPAHEGL	VATVKE	AGRSI	HEIPRE	LHRTPEL	PLA-----	PRP-LKE	GSIT	1452				
Dy	1061	LAPPL	LSP-----	LLPSIK	PTVRKTRQ	DGSGQE-----	APEAP	LSSELEFP	HKPIR	TRKSS	1113				
Qy	1453	QGTPLK	YDTGASITG	SKKHVRS	LIGSP	GRTFPV	PHPLDV	MADARALER	ACYESL	KSRP	1512				
Dy	1114	RMTFP	-----	ATSAPE	PHFSTS-----	TAQ	PVTPKPT-----	-----	SOATSR	-----	1149				
Qy	1513	GTASSG	SGSIARG	APV1-----	VP	ELGK	PROSLTYED	HGAPAGHL	PRGSPVT	MEPTRELQ	1570				
Dy	1150	-----	TNRSSVKT	PEPVPTAP	EL-----	QST	STSD-----	-----	OPVT	SEPTSQVT	1187				
Qy	1571	EGSLSS	-----	-----	KASQ	RKLTS	TPREIA-----	KSPHST	YPERH	PHIP	1610				
Dy	1188	RGRKSR	SVKTPETV	VTVALE	LQBSTD	TRDPVT	SEPTSQAT	RKGRNSVKT	-----	PEPVVP	1245				
Qy	1611	YELLRG	VGVDL	YRSHI	PLAFDPT-----	-----	SIP	GIPLD	AAAAAYLP-----	RH	1654				
Dy	1246	TAPEL	QSTSDQ-----	PVTSE	PTQATGR	KNRSSVKT	PEPVPTAP	ELRPFST	STDRP	1300					
Qy	1655	LAPNPT	YPLVPP	ILRGYP	DTALE	ENKOTI	INDYITS	QOMHNT	TATAMA	QADMLRGLS	1714				
Dy	1301	VT	PKPT-----	SR	TTSRTNM	SVKTPETVVP-----	TAPEL	QISTSD-----	-----	OPVT	1343				
Qy	1715	PRESS	LALYAAG	PRGI	IDL	ISOV	PHLVLV	VPPTFG	-----	TPATAM	DRLAYL	TAPOPF	-----	1768	
Dy	1344	PKPTS	RT-----	TR	SRTNM	SVKNP	ESTVPIA	ELPPST	STEQ-----	-----	PVTPE	TSRATR	1393		
Qy	1769	-----	SRHSS	PLSP	CGP	PHLTKPTT	TSSE	RRDRDRDR	REKISIL	ITSTT	VEH	1823			
Dy	1394	GRKNR	SGKTPET	LVP	TAPEK-----	LEP	STSDQ	PVTPEPT	SOATGR	TRNRSSVKT	PETVVPT	1451			
Qy	1824	APIWR	P-----	GTEOSS	SGSSG	GGG	SSSR-----	-----	-----	PASH	SHAHQ	HSPI	-----	S	1864
Dy	1452	AP	ELQ	STSDQ	PVTPEPT	SOATGR	TRDRSSVKT	PETVVPTAP	ELQ	AS	STDO	QVTS	SDPT	1511	
Qy	1865	PR	QDALQ	QRP	SVLHNT	CMKGII	IT	AVBFS	KPTVL	LRSTSTSP	VRPAATFP	PATH	CPGLG	-----	1922

Db 58 RMPDCSVALPPFSISKQHAIEIILAWDKAPILRDCGS-LNGTQILRP-----PKVLSPGVS 112
Qy 75 GNERSOLE-----HLRPSHSY-----LPELGKSEMEFIESKPRLELLPDPILRSPILLATG 127
Db 113 HRLRQDELILFADLLCQVHRLDVLSPFVSRGPLTVEETPRVQGETQORLL----- 163
Qy 128 QPASEDLTKDRSLTGKLEPVSPSPHPTDPELSELPRLSKBELIQNMDRVDREITMVE 187
Db 164 -----LAED-----SEEVFLSER----- 178
Qy 188 QOISKLKKKQOOLEEAAKPEPEKPVSPPPPIESKHSRLVQI1YDENRKAEAHRILEG 247
Db 179 -----RWKKSSTSSSVIPESDECHSP-----VLGG 207
Qy 248 LGPOVELPLYNOPSTROYHENIKINQAMRKLILYFKRNRHARKQWKQFCQRYDQIME 307
Db 208 LGPPF-----AFNLNSDT-----DVBE 224
Qy 308 ALEKVERIENPR-----RAKESKVREYKOPFEIRKQRELQERMQRSGVQSGG----- 360
Db 225 GQOPATEAASAARGATVEAKQSEAKVVTIQL-----KQPLVKERDNDTKVKGAGNGV 282
Qy 361 -----LMSAARSEHEVSEI1D-----GLSEQENLEKQMLQAVIPMLY-----DAQ 404
Db 283 VPAGVILERSQPPGSDSDTDVDDSRPPGRPAEVHLERAQ-----PFGFIDSDDTAAE 335
Qy 405 ORIKINNGLMADPMKY-YKDRQVNN-----MWSEOEKETPREKFMQHPKNFG 452
Db 336 ERIP-----ATPVVPMKRRKIFHGVGTGRGAPGLAHLQSSQAGSDTDVEEGKAPQ 387
Qy 453 LIASFLEKRTVAECVLYVYLTAKKENYKSLVRRSYRRRGKSQOQOQOQOQOQOQOQPM 512
Db 388 AVP-----LE-KSQASWINSDTDDEEVSAA1LAHLK-----ESQPA1WNRDAEDM 435
Qy 513 PRS-----SQEKEDEKEKEAEKEKEKEVENDKEDLLK-----EKTDTSGE 556
Db 436 PQRVLLQORSQTTTERSDTDVEEBELPVEN-REAVLKDHKTIRALVRAHSEKQDPPFGD 494
Qy 557 DNDEKEA-VASGKRTANSQGRKGRITRSMANEANSEAI-----TPQOS-- 601
Db 495 SDDSVEADKSSPGIHLERSQASTVDINTQVEKEVPPGSAI1HKKHOVSVEGTNQTDKV 554
Qy 602 -----AELASME1NESSRW-----TEEMETA-----KKGLL-----EHGRNWSA 636
Db 555 AVGGPAKLLVLSLEA-WPLHGDCE1DAEGTALTASVAVADVRSQSLPAEGDAGAEEWA 612
Qy 637 I-----ARMVGSK-----TVSQCKNFYNYKKQNLDLE1LOQHLKWEKERNARRKKKAP 687
Db 613 AVLKQERAEVGAQGGPPVAQVEQDL--PISREN1DLVVD1DTL1GESTQPOR----- 663
Qy 688 AAASEEAAFPVVEDEMEASGV-----GNEEEMVEEA-----EALHASGNEV 731
Db 664 -----EGAQVPTGREHQVGTGKUSEDNYGSDLDLQATOCFLENQGLEAVQSWEDEP 718
Qy 732 PR-----GRC-----GPATVNNSSDTESTIP-SPHTEAAKDT 762
Db 719 TQAFMLTPPQELGPHSCHSFQTTGLTDEPWEVLATQFCLRESESDSETOPFDTHLEA----- 774
Qy 763 QONGP-KPATL1GADGPPGP-----PTPRTSRAP1E-----PTPASEATGATPPPA 811
Db 775 --YGPCLSPRA1PGDHPESPVHTEPMGIQGRQTVDKWGI1PKETAERVG-----PER 828
Qy 812 PRSPSPAPPPVW-----KEEKEEETAAAPPE-----EGEEQKPPA-AB 849
Db 829 GPLERETKLLPERQTVTGEELTKGQDREQKQLLARDTQOESDKNGESASPERDRE 888
Qy 850 ELAVDTGKAE-----PVKSECTEBAEAGPAKGDAEAB-----ATAEGALKA 893
Db 889 SLAKVETSE1QEKVQKQTLPSKAFEREVEPVANRECDPAELKEEVPK1VILERD1QR 948
Qy 894 EKEGGSGRATTAKSGAPD-----SDSSATCSADEVBAEGDKNRL1SPR-----PSLL 945
Db 949 GEPEGSGDQKQASSPTPEPGVGAGDLPGFTSAPVPSGSGGGRGSPVSPRRHKGSGLL 1008

Qy 946 TPTGDPANASQKPLDLKQKQRAA1PPIQVTKV-----HEPPREDAAPTKPAAPP- 1000
Db 1009 NCKMPFAEKAS-----R1RAA-----EKVSRGDQSPDACLPPTVPEAPAPQ 1051
Qy 1001 PPONLQESDAPQOQSGSPRGKSRGPAPPADKEAFAAEAKQLPGDPPCWTSGL-PF-PVP 1058
Db 1052 KPLNSQSKHLAPPLP1LLPSIKPTVRKTRQDGSQEAPEAP-----LSSELEPHPKP 1106
Qy 1059 PREVIKASHADPPSAFSAVAPCHPLPLGLH1H1D1ARVLP1RP-----PT1SNPP 1107
Db 1107 K1RTRKSSMTFPP-ATSAAPEHPBT-----STAQPVTKPTSQATRSRTNRSVKTPEP 1161
Qy 1108 L1SSAKHPSVLERQ1GAI1SQGMSVQLHVPFSEHAKAPGVPTMGLPLPMDPKKLAPFSGV 1167
Db 1162 VVPTA-----FELQPS1TSDQPVTS-----EPTSQV 1187
Qy 1168 KQQLSPRGQAGPPELSGVPTAQEASVLRGTALSGVPGSITK1G1P1ST1RVPSDSALT1RG 1227
Db 1188 TRGRKS-RSSVKTPETV-VPTALEQ-----PST----- 1214
Qy 1228 S1THGTPADVLYKGT1TRIIGEDSPSRDLGRGDSLPKGHV1YEGKKGHVLSYEGGMSVT 1287
Db 1215 -----STD1RPVTSEPTSQATGRKN-----RSSVKTEP-----PVV 1245
Qy 1288 QCSKEDGRSSSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAI1P1RHS1PHLKE 1347
Db 1246 PTAPELQPS1TSDQPVTSBP--TYQATRGKRNRS-SVKTPEPVPTA--PE----- 1291
Qy 1348 QH1RGSITOG1PRSVEAQEDVLRK1KKEEGTPPPPPSRDLTEAVK1QALG1PLK 1407
Db 1292 ---LRP1STSD1RPVT-----FKP1TSR---TT1RSRTN----- 1316
Qy 1408 KPAHEGLVATVKEAGRS1HE1PREELRHTPELAPR1KEGSITQ1G1P1KYDTGASTTG 1467
Db 1317 -----MSSVK1TPE1VPTAPELQ1STSDQ1VTKPT-----SRTTR 1353
Qy 1468 SKH1D1RSL1GSPGRT1PPVH1P1DV1WADARALERAC1VEESLKS1RPGTASSGSG1IARGAP 1527
Db 1354 SRTN--MSSVKNPE1STVP----- 1369
Qy 1528 V1VPELCKPQSP1TY1EDH1GAPFAGH1PGRGSPV1TMEPT1PRLQ1EGLSS1SSKASQDRK1TS 1587
Db 1370 -IAP1EL-----P1E1T1ST1E-----QPV1T--PEPT-----SRA1TGRK-- 1397
Qy 1588 TP1R1AKSPH1STVP1EH1PH1P1SPYEH1LLRGVSGVD1LYRSH1P1AFD1D1TS1P1RG1D1DAA 1647
Db 1398 -NRSSGKTPE1TV-----TAPK--LEPST 1419
Qy 1648 AY1LPRH1ANPTVPH1LYPP1LYRG1PDTAA1ENRQ1I1NDY1ITSQOMH1NTATAMAQRA 1707
Db 1420 STDQF--VTPEPT-----SQATRG1TRNS1SVK1TPE1TVP--TAPELQ1F1ST1D-- 1463
Qy 1708 DMLRGLSPRESS1ALNVAAG1RGI1D1LSQVPH1PVLVPT1PGT1PATMDR1LAY1LPTAQ1P 1767
Db 1464 ---QPV1TPE1T1SQA-----TRG1TD1RS1VK-----TPE1TV-----VETAP1EL 1497
Qy 1768 FSSRHSS1SP1SGP1G1TH1K1TPT1TSS1SER1DRDRDRDREREK1S1LT1T1TVEHAP1W 1827
Db 1498 QASAS1TDQ1PVT-----SDPT1SRT-----RGRKNRS1SVK1TPE1TVVPAAP1EL 1538
Qy 1828 RP--GTEQSSGSGSGSGGSSSRPASHAHQHS1P1SR1TQD1ALQORP1SVLHNTGMKG 1885
Db 1539 QP1P1TSDQ1FV1TPE1TSR1ATRG1RTNR-SVKTPE1SV1P1AP1ELQ-----P1TSR1NQ-- 1587
Qy 1886 I1T1AVEP1SKPT1VLR-----ST1SSP1VRPAAT1FP1ATH1CPLG1GLD1GV1T1P1ME1V1LP1KE 1941
Db 1588 LV1TPE1TSR1ATRC1TRNS1SVK1TPE1V1VPTAPE1PHPT-----T1SDQ1PVT1PK 1634
Qy 1942 APRVARPERPAD1TCHAF1LAP--PARSGLEPAS1PSK1SGEP1R1P1V1PVS1GHAT1ARTPA 1999
Db 1635 TS1R1TRK1TRNS1SVK--TPK1PVEP1AASD1LE1FT1DQ1SVTPE1A1AQ--GQ1SK1LRS1ST 1689

QY 2000 KNLPHASDPDPAPPASASDPRHREKTQSKPFSIQEILRLSLGYGSSYSPGCVFVSPV 2059
 Db 1690 VRAMPVPTTPEFQSP-----VTTDQPI-----SPS-----PI 1716
 QY 2060 SPSLTHDKGLPKHLELDKSHLEGELRPQGPVKLGGEAAHPLRLPL-PESOPSSSP 2118
 Db 1717 TOPSC-----TKRQRAAG-----NPGSLAIPIDHKCSAPLEPKSQASRN- 1756
 QY 2119 LQTAGVKGHORVVTLAQIHISEVITODYTRHHPOOLSAPLPAPLYSPFGASCPVLDLRR 2178
 Db 1757 --QRGWAVRAESLTAIPASPQOLLE--TPHSAQIQKVEPAGRSRFTPELPQKASQSR 1812
 QY 2179 PPSDLYLPPPHGAPARGSPHSGKRSPEPNKTSVLGGEGDIEFVPPGEGMTPEGHSR 2238
 Db 1813 KRLSLATMSP-----PHOKQQRGEVSKQTVIIRKEE-----DTAEPKGRKEE 1855
 QY 2239 SAVYF--LLYRDGQTPSRMGSKSPGNTSQPPAPFSSKLTESNSA 2281
 Db 1856 DVVTFPKGKKRQDAEEENRIPSLRRT-----KLNGESTA 1893
 RESULT 79
 Q9VWC0 PRELIMINARY; PRT; 2529 AA.
 AC Q9VWC0;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE CG32529 protein.
 GN CG32529 OR CG11936 OR CG15619.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doult L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Letitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milbeh N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter C., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Gibson J.A., An H., Baldwin D., Banton J., Beeson K.Y., Dietz S.M.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Buetz S.M.,
 RA Dodson K., Dorsett V., Doult L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrieria S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle B.J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003513; AAP49026.2;
 DR FlyBase; FBGN0052529; CG32529.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR InterPro; IPR001025; BAH.
 DR SMART; SM00439; BAH; 1.
 SQ SEQUENCE 2529 AA; 270675 MW; C93B1726EE7547D4 CRC64;
 Query Match 3.2%; Score 421; DB 5; Length 2529;
 Best Local Similarity 18.5%; Pred. No. 1.2e-11;
 Matches 463; Conservative 287; Mismatches 937; Indels 816; Gaps 111;
 QY 472 LTKKENYKSLVRS-----YRRGKSQQQQQQQQQQQQQQQQQQQQPMSRQEEKEK 525
 Db 2 VSKRKNAAERVQRSTATAAAAAAATASTQGGKRRPGKQTAKSGGGGDEADTKKS 61
 QY 526 EKEAEKEEKEPEVNDKEDLLKEKTTDDTSGDNDKEAVASKGRKTNQSQ---GRKGR 582
 Db 62 ASKTDSAARKP-----AKOPTSGSGVSTATTEGAGNGGKGQ 98
 QY 583 TRSMANESAEITPOQSABLA---SMELNESSRWTEEMETAKKGLLE-----HG--- 631
 Db 99 TTSSTTSSTTS-STTPASASASASNSKLSPERHEDGKARAINKMLKSLDIKHGPDN 157
 QY 632 -----RWNSAIARMVSKTYSOCKNFYFNKKR----- 659
 Db 158 LLPSGEERLNDGVLSKISENVKTKSRAAAVKVSSLNPGKMPVPRVPSPEPVREKVA 217
 QY 660 -QNLDLITQOHLKWEKERNARRKKKAPAA-----ASEEAAFPVVEDEEAAA----- 707
 Db 218 EXPKAEKVLDRKPKAEQIITPPAPESPASKVPKKTVOQAKKAVESLIVQATVLTTS 277
 QY 708 -----SGVSGNEEEMVEAEALHASGNEVPRGCSGPATV-----NNSSDTESIPSP 754

278	DB	AATSAPSSASTSKHTSYNPKGKSAHSEQMPKELSKLAKMGKPKKKEANLVKSTARLPDS	337
755	QY	HTIAAK-----DTQNGPKPATLIGADPPPGPTTPRRTSRAPTEPTPASEATGAPT	807
338	DB	NTESVQMELEBIVSTPIVTPSAT-----PTATPT-----TATPTLTPTTATRTTPT	388
808	QY	P-----PPAPSPSAPPVVPKEE-----KEETAAAP--VEEGEE	842
389	DB	PTLTPTPTSTPSSTAIAAGQAPPPP--PPGAPKQRPVKYKTKTSAGSKRMOQALN	446
843	QY	OKPAA-----ELAVDTGKAESPVK-----SCT--	867
447	DB	QKAPASWSGAKIYDFKSGSEDEPIKMWLPSELREFSDEDNKLFLERQASBTKV	506
868	QY	--BEAEGPAKGDAEAAETAEGALKAEEKGSGR-----ATTAKSGAPQDSSAT	920
507	DB	QVQDEEQPLVKE-EXPPSEKVKVKSKEKEKEDVDPLSNAQSPAPAAKHTTST	565
921	QY	CSADEVDEAEGGDKNRLSPRSL-----TPTCDPRNAPSQKPLD	962
566	DB	TSITSIYSVKQAKRKIATPMPSSAGRKKTGKGAAAKKEPPSGSSAESSEDERK	625
963	QY	LKQL-----KORAAAIPPI-----QVTKVH-----PPR-----	986
626	DB	LSAYSGPKRHMASNALAKVQCLYENESHTAHGLSKATQOPIRLTLDGDDDNHKD	685
987	QY	-----EAAAPTK-----APPAP--PPONLOPESDAPQOPGSSPRKSRSPAPPAD	1031
686	DB	SPRHGDTDRDKGVSVPHTSAAAPPAAPAPKAEAPRELYVPGRGVKGWE-----FD	741
1032	QY	KEAPAAEAKQLPGDPCCWTSGLPFPVPPREVIKASPHADPFSAFSAPPGHPLPLGLHDT	1091
742	DB	SDSETEELQLOTLPPVKKKLPKKAPKVASSETDAKRKKKSQVVEES-----DDE	794
1092	QY	ARVLPRPPTISNDPPLISSAKHPSVULEROIG-----AISQMSVOL	1133
795	DB	EAKEKEKEKPKPKPKPKKRTAFTEELIGDYKGLARKRWASLNASAIVAATYEVR	854
1134	QY	HV-----PYSEHAKAPVGPVTMG-----LPLMPDKKLAPFGSVQEOAL	1172
855	DB	HLDKNKLASDCSPSDELTMTPTTSKVTSAAAKTEVHIHKKPKB-----SKESL	907
1173	QY	SPRQAGPPBSLGVPTAQEASVLRGATLGSVPGSITKGPSTRPVSDSAITYRGSITHG	1232
908	DB	RERERESER-----RRERONELKDSVGMKSSHIIESNDSKYSKE-----TKE	954
1233	QY	TPADVLYGTITRIIGDSPSRLDRGDSLSLPGHVIYEGKGHVLSYEGCMSVTQCSKE	1292
955	DB	TNTDT---TTTTGYRDSAASTKADKDCSRPTSSV-----VIVQDTDVIT---	998
1293	QY	DG---RSSGPPHETAAPKTYDMWEGV-----GRAISSASTE-----GLMGRAI	1335
999	DB	-GVYVNASG-----AAQEAYCKQYRVQSSVTEERVLRPGSVPEPKSVTPLSALSSML	1051
1336	QY	PP-----ERHS-----PHLKEQHIRGSITQGPISRYVEAQEDYLREAK	1376
1052	DB	PPGASSGLSEASHSPPIPVHNAPTGPHVLPGEHLSAGM-----YHAPDLGLTHEA	1103
1377	QY	LLKREGTPPPPPPS-----RDLEAYKTOALGP-----LKUKPAHEGLVATVKEAGRS	1424
1104	DB	LPDHHG---PPPSVAAAAAAYHAHOLAPGGGVGLHMHQHQ-----YAAHAHA	1156
1425	QY	IHEIPRELRHTPELPL-----APRPLEKGSITQGTPLKYDTGASTTCKSKHDRVSLIGSP	1480
1157	DB	HHYQQAEBYHGGPPPPPHHYGGPPPOGHGPPGPVFPVAVENTVTVFPE-RCDVGSP	1515
1481	QY	GRTPFPVHPLDVMADARALACRYEESLSRPGTA-----SSSGSGIARGAPVIV---	1530
1216	DB	-----PP-----SYRASAYPSPSSVGMPPPTLGGSSAFACAPSLHHQH	1255
1531	QY	-----PELGKPROSPLTVEDHGAFFAGHLPRGSPVTWREPTP-----RLQE-----	1571
1256	DB	OHQOPGCVGYPHHDPGSGYOPAGFLI-----SPHVIAPPPPAQVKLQEDGVPVSA	1309

[illegible]

```
RESULT 80
Q14676
ID Q14676 PRELIMINARY; PRT; 2089 AA.
AC Q14676;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein KIAA0170.
GN KIAA0170.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid leukemia cells;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
DR EMBL; D79992; BA011487.1; -.
DR Genbank; HGNC:21163; MDC1.
DR GO; GO:0005622; C:intracellular; IEA.
DR InterPro; IPR001357; BRCT.
DR InterPro; IPR000253; FHA.
DR InterPro; IPR008984; SMAD_FHA.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00292; BRCT; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS00172; BRCT; 1.
DR PROSITE; PS00006; FHA_DOMAIN; 1.
KW Hypothetical protein.
KW SEQUENCE 2089 AA; 226690 MM; A628EC7567E850 CRC64;

Query Match 3.2%; Score 419.5; DB 4; Length 2089;
Best Local Similarity 19.6%; Pred. No. 1.1e-11;
Matches 425; Conservative 256; Mismatches 762; Indels 725; Gaps 106;

Qy 486 SYRRRGKQQQQQQQQQ---QQQQQQQPMRPSQOEKDEKEKEAE-----KE 532
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 235 SAARRGVAEAKQAEAVVTVLEKQDQPLVK---ERDNDTKVKRGAGNGVWVAGVILE 290
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 533 EEKPEVNDKEDLLKEKTDTS-----GENDKE-----AV 564
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 291 RSQPPGEDSDTDV-----DDSRPPGRPAEVLHRAQPFPGFIDSDTDAEERIIPATPVI 345
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 565 ASKGRKTANSQGR-----KGRITRMA-NEANSEEAITPQQ---SAELASMEINSSR 614
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 346 PMKKRKIPHGVTGRCGAPGLAHLOESQAGSDTVEECKAPQAVPLEKSQASWINSND-- 403
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 615 WTEBEMETAKGLLEHGRN-----WS-----AIARMVGSKTVSOCKNPFYNYKKRQ 660
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 404 -TDDEEVSAALTLAHLKESQPAIWNRAEDMPQVVLQORSQTTE-----R 451
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 661 NLDEILOQHLKMEKERNARK---KKKAPAAASEEAFFPVVE--DEEMEAASVSGN--- 713
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 452 DSDTDVEEELPVE-NREAVLKDHTKIRALVRAHSEKQDQPPFGSDSDSVEADKSPGIHL 510
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 714 -----EEMVBEAEALHASGNEVPRGCGPATVNNSDTESIPSP----- 754
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 511 ERSQASTTVDINTQVEKEVPPGSAIMHIKKHV-----SVEGTNQTDVKAQGGPAKLL 563
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 755 -----HTEAAKDTGQNGKPPATLGADGPPGPPPTPRRTSRAPISPTASEATG 804
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 564 VVSLAEAWPLHGDCEQD-ABEGTSLTSASWAD-----VRKSQLPAGCDAGAEWAA 612
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 805 APTPPAPPSPAPPPVVPKEKEEETAA--APPVEEGEOKPPAAE---FLAVDTGKAE 859
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 613 A-----VLKQERAHEVGAQGGPPVAQVEQDLPISRENLTDLVVDV----- 652
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Qy 860 EPVKSECTEAEAGPAKGKDAEAAEATAGALKAEKKEGGSGRATTAKSGAPQSDSSA 919
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 653 -DTLGESTQPOREG-----AQVPTGREQHVGGTKDSEDNYGDSEDLDLQA 698
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 920 T-CSADEVDEAEAGDKNRLSLSPRS---LLTPTGDPRANASPOKPLDLKQLKQRAAIP 975
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 699 TQCFL---ENQGLEAVQSMEDPTQAFMLTP-----PQE-----LGP 732
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 976 ---IQVTKVHEPPREDAAPTTPAPPQPPQNLQPSDAPQQQSGSSPRGKSRSAPPAD 1031
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 733 SHCSFQTTGTLDEPWEVLA-----TQPFCLRESEDETQPPDT-----H 771
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1032 KEAFAPAEQAQKLPDPGPPCWTSGLPPVPREVIKASPHADPDSAFSAPGPHLPGLG 1091
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 772 LEAYG-----PCLIS-----AIPGDHPES-----PVHTEPMGIQGR 805
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1092 ARPVLPRPTTISNPPPLISSAKHPSVLERQIGALSQMSVOLHVPYSEHAKAPCV 1151
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 806 GR-----QTVDKVMGIP--KETAEVGP-ERG 829
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1152 LPLPMDPKKLAP-----FSGV-----KQQLSPR-----GQAGPPE--- 1182
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 830 -PLEREYEKLLPERQTDVTGEEELTKGQDREQQLLARDTQOESDKNGESASPERDRE 888
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1183 --SLGVPTAQEASVLRGTALGSVFGSITKGIPTSTRVPSDSAITVRSITHTGTAD 1240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 889 SLKVEIETSEBIQ-----EKQVQKTLPSKA-----FE 916
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1241 GTITRIIG--EDSPRLDRGREDLSLPKHVIEGKGHVLSEYEGM-----SV 1293
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 917 REVERPVANRECDPAEL---EEKVPKVLBERDTQOGE---EGSQDQKQOASPTPEP 969
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1294 GRSSGGPPHETAAPKRTYDMMEGRVGRAISSASTEGLMGRAIPPERHSHPHLKEQH 1353
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 970 GVGAGDLPGTISAP-----VPSGSGSGRGSPVSPRRHQ-----KG 1005
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1354 SITQIP-----RSYVBEAQEDYLREKLLKREG---TTP-----PPPSRD 1400
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1006 LLNCKMPPAEKASIRAAEKVSRGD---QESPDACLPPAVPEAPAPPPKPLNSQSK 1060
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1401 ALGPLKPAHEGLVATVKEAGRSIHIEIPRELBHTTPELPLA-----PRP-LKEG 1452
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1061 LAPPPLLSPP---LLPSIKPTVRKTQDQSGOE---APEAPLSSELEPPHPKPKIN 1113
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1453 QGTPLKDYDTGASTTGSKKHVRSLSIGSPGRTFPPVPHLDVMDARALERACVEESL 1512
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1114 RMTFPP---ATSAPEHPST-----TAQVTPKPT-----SQATRSR- 1149
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1513 GTASSSGGSIARGAPVI--VPELQKPRQSPITYEDHGAPFAGHLPGRGSPVTMRB 1570
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1150 ---TNRSSVKTPPEVVPVTAPEL---QPSTSTD-----QPVT-SEPTSQVT 1187
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1571 EGSLSSSKASQDKLTSTPREIAKSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSH 1630
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1188 RGRKRSRSVKTPETVWPTALELQESTST---DRPVTSEPTSQATRGKRNSSVKTP 1243
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1631 AFDPSTIPRGIPLDAAAYLPHRLAPNPTVPHLYPPVILRGYPDPTALENKRQTI 1690
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1244 ---VTAPELQESTSTDOP--VTSEPTY-----QATGRKNRSVKTPPEVVP--- 1286
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1691 TSQOMHNTATAMAQADMLRGLSPRESSLALNAAQGRGIIDLSQVPHLPVLVPPT 1750
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1287 TAPELRPSSTD-----RPVTPKPTSRT-----TRSRTNMSSVK-----T 1321
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1751 PATAMDRLAYLPTAQPFSSHRSSPLSPGPGTHLTKPTTTSSSERERDRDRDRDR 1810
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1322 PETV-----VPTAPELQISTSDQPVTP-----KPTSETT-----RSRTN 1356
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1811 EKSLTSTTTTVEHAPIWRP--GTEQ-----SSGSSGSSGGGGSS-----SR 1855
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1357 MSSVKNPESTVPIAPELPPSTSTSEQVTPPETSRATGRKNRSKGTPTETLVPT 1416
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1856 HAHQHSPIPR-TQDALQQRPSVLHNTGMKGIIT--AVEPSKPTVLRSTSTSSPV 1912
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 1417 STSDQVTEPTSQATGR-----TNRSSVKTPTVVTAPBLOPSTSDQVTEPT 1470
Qy 1913 F-----PPATHCLGCTLDGVVPTLMPEVLLKPEARVPRPRADTGHAF 1959
Db 1471 SQATRGTRDSSVKTPTVVTAPBLOQASSTDPVTEPTTRTRGRKRRSSVKTPTV 1530
Qy 1960 LAKPPARSGLEPASPSKGSBPRPLVPPVSGHATIAETPAKNLAPHASDPDPAPPASAS 2019
Db 1531 V---PAPELOPPTSTD-----RPTPTPTSTRATGRNRSV---KTPESIVPIAPEL 1578
Qy 2020 DPHEKTKQSPFSTOELESLGSLVHGSSYSPGVEPVSPVSSPSTLHDKGLPHLELDK 2079
Db 1579 QPSTSRNLQVTEPTSTRATCRTRNRSSVKTPTVVTAPBLOPSTSDQVTEPTKLTGRAT 1638
Qy 2080 SHLEGEALPKOPGVKLGGERAHLPLRLPLES-----QPSSSPLLOTA-----PG 2125
Db 1639 RRKTRNRSSVKTPTKPEV---PAASDLEPFTTQDQVTPBIAAQGGOKTLRSSTVRAMPVPT 1696
Qy 2126 VKGQVVTLLAQHIS-EVITO-----DYTRHHQQLSAPLPAPLYSFP----- 2167
Db 1697 TPEQSPVTTQDQVTPBIAAQGGOKTLRSSTVRAMPVPT----- 2167
Qy 2168 --GASCPVLDLRRPPSDLYLPPPHGAPARGSPHSGKSPENKTSVLGGGEGDIEPV 2225
Db 1757 RWGA-----VRAAESLTAIDEP-----ASPQLLETPIHASQIQKVEPA 1794
Qy 2226 SPPEGMTEPHGRANVPLLYDGEOTEPEPBGWS-KSPGNTSQPPAPFSLKTESAMVK 2284
Db 1795 -----GRSRFTPELOPKASQSKRLATMDSPPHOKOP----- 1827
Qy 2285 SKQOEINKLTHNRNEPEYNISQGTIEFNNPAITGTGLMYRSQAVQEHASTNMGLEA 2344
Db 1828 -QREVSQK-TVIIKEEEDTAEPKKE-----EDV 1856
Qy 2345 IIRKALMGKYDOWEESP---PLSANAFNPLNASLPAAMPITAADGRSDHLLTSPG-- 2399
Db 1857 VTPKPKRQDQAEENRIPSRSLRKLQNESTAPKVLFTGVVDARGERAVLGGSL 1916
Qy 2400 -GGKAKVS 2406
Db 1917 AGSAAEAS 1924

RESULT 81
ID Q17343 PRELIMINARY; PRT; 6994 AA.
AC Q17343;
DC Q17343;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE UNC-44 ankyrins.
GN UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=95263663; PubMed=7744957;
RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,
RA Boontrakulpoontawe P., Jeyaprasath A., Hedgecock E., Wheaton V.I.,
RA Sobery A.;
RT "An ankyrin-related gene (unc-44) is necessary for proper axonal
RT guidance in Caenorhabditis elegans.";
RL J. Cell Biol. 129:1081-1092(1995).
RN [2]
RP REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.
RC STRAIN=N2;
RA Otsuka A.J.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39847; AAB41827.1; -.
```

```
DR EMBL; U21733; AAB38384.1; -.
DR HSP; P42773; 11HB.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR000906; ZU5.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00791; ZUS; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 21.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00218; ZUS; 1.
DR PROSITE; PS50088; ANK_REPEAT; 22.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 2.
DR ANK repeat; Repeat.
SQ SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;

Query Match 3.2%; Score 419; DB 5; Length 6994;
Best Local Similarity 17.5%; Pred. No. 5.5e-11;
Matches 516; Conservative 371; Mismatches 1020; Indels 1040; Gaps 129;

Qy 16 ERYPP-----HSLSPV-----QIARTHTDVLGLVYQHHSRDYASHL--SPG 56
Db 3169 DPFSAPSHSVSHGSHASESPVAQQEIPQTRFEHDSFAAQY-FHEEYEVSHLVEQA 3227
Qy 57 SIIQPPRRP-----
Db 3228 PLUTEQQHQFSGDESGELGSKVLGFAKAGVAGVVAAPVALAAVGAAYDALKK 3287
Qy 71 EFQGNERSQELHLRPESHSLPELGKSEMEFTESKPRLELLPDLLR----- 119
Db 3288 DDDEEDQERESLLQERSIDSFH--ASEQSIIEEHERFEESFPVSEKHHVTETTTTTT 3345
Qy 120 -----PSPLLA-----TGOPAGSE-----DLTKDRSLTKLEVPSPSPPTDP 158
Db 3346 VTREYFDEHPLVSOIEGEGKSPASEKSLPHVVETTTTTTTRFDKNDSESPVSEK 3405
Qy 159 EELVPPRLSKBELIQNMDRVREITWVEQOI SKLKKQQQLSEEAAPPEPEKPVSPPP 218
Db 3406 EQE---PTTVSRVYETAEGDEPEHHVTETTTTTTKEVIDDSDQEMGDDDEP-KQESPOV 3461
Qy 219 IESKRSVLQIIVDEN---RKAEEAARI-----
Db 3462 VETTTTTTSREYDNDDETREAGDSHITETTKTTTVTRFHFGEQPEETEETDEVELP 3521
Qy 250 POVE---LPLYNOPSD-----TROVHENIKINQAMKLLILY 283
Db 3522 PKTEEDNVSEYSESSTSVSRVPRDPDEPHILIEITTTTTTVTRVHNEPEETVDDQDAPI 3581
Qy 284 FKRNHARKQWKQFCQRYDQLMEALKKVERIENRRRAKSKVREYKEKQPEPR-- 341
Db 3582 SFSQEH-QDDDSQASHQHDR--ESPVESEKSVKHTTTTTTTTTTQLYDDEASEIRGE 3638
Qy 342 -----KQRELQERMSRVGSGS 359
Db 3639 SPVATESEHVSSTKSDSESHQVPSVIETTTTTTVTRFYDDQDELQREDHTOSEBRS 3698
Qy 360 GLMSAARSSEHVESEII-----DGLSEQENLEK-QMRQLAVIP-----P 397
Db 3699 --SIPTETHEHSHLKEITTTTTTVTRFYDPENVKQLQDSQFSLSPSHVESEIYVP 3756
Qy 398 MLYDADQORI-----KFINNGLMADPMKVYKDRQVNMWMSOEKE 438
Db 3757 ESPVAKQOEIPQTRFEHDSFAAQYFHEDEYEHQVTEQAPLITTEQHQH-PESGEESDGE 3815
Qy 439 TFRKFMQHPKNGFLIASFLERKTVAECSVL-----YYLTKN-----ENVKSLVR--RS 486
Db EMBL; U39847; AAB41827.1; -.
```


Db 4287 GQGESDQIPSEISITSENMDRETSSSPVQNRDEEFVLPALPYKQTEYGRVDSDHAP 4346
Qy 811 APPSPAPPPVVPKEKEEETAAAPPVEGEE-----QKPPAAEELAVDTGKAEEPVKS 864
Db 4347 ASPAPSAESPI--QAYKQESQEAHSLQFOQRSSVSHEDSPAQYFHDNEDHDSVPVS 4404
Qy 865 E---CTEAEAGPAGKDAEAAEATAEALKAKEKG--GSGRATTAKSSGAPQDSQSAT 920
Db 4405 DRAPLLTQQHQCPESGEED--GEGFGSKVLGFAKKAAGVAGVVAAPVALAAVGAAYD 4463
Qy 921 CSADVDVAEGGDKNRLSPRLSLTPTGDRANASPKPLD--LKQLKQRAAAIPPIQ-- 977
Db 4464 ALKKDDDEDNDPEQKLPK-----SPERQVLNVPEDSEISELEYTSPSPTEK 4517
Qy 978 -----VTKVHPPPRED-----APTAPPAP----- 999
Db 4518 SESQCYTETVTTVTREYLDLPQSVTRSPSEHDISEQYAPESPVQDPVVEKITTV 4577
Qy 1000 -----PPQNLQ-----PESAPQPGSSPRG-----KSR 1024
Db 4578 IRQVHDEPPQIEEOTIPEEVTVLREVYESPEGDEPEQHYIETKTITITKEVHVVPVEED 4637
Qy 1025 SPAPADKEAFAAQAQLPGPPCWTSGLPVPVPREVKA-----SPHAPDPSAFY 1077
Db 4638 VOISPVHSETSVSEKQ--LPAD-----EQLDFEVTESTTTATVTRYEYEEVPPPGSE 4691
Qy 1078 APPGHPLGLGHDARTVLPRLPTIS-----NPPPLISSAKHPSVLEROI 1122
Db 4692 ADDESHAKYMETTTTTVTREYEVSEDEDEHQSOQVORDSPAPSESVKH--VIEK-- 4747
Qy 1123 GAISQMSVQLVVPVSEHAKAP-----VGPVTMGLPLMDP-----K 1159
Db 4748 -TTTTVTREYEPEDSHSPVEDDVGFKVTITTTTTVTTHPEPEPPDEHVVSE 4806
Qy 1160 KLAPEGVKQLSPRG-----QAGPPESLGVPTAQEASVL 1195
Db 4807 RYAGSPVPSEDSREIETTTTTVTREHFELEDQDHVVSEYQSASGSPVPSEKVE 4866
Qy 1196 RGTALGVPFGSITK-----GIP-----STRVPSDAITYRGSITHTGPADV 1237
Db 4867 R--VIETTTVTVTRHEHEDDIPTIVETSHDDPAASVSEEDV-----HQIQT 4917
Qy 1238 LYKGITR--IIGEDSPRLDRGSDSLPKGHVIEYEGKGVLSYEGMSV-----TQC 1289
Db 4918 TTTTIVTRHVVPDE--EIDSGRMDELEK-----YSSSPVPSEDSRVETITTTTTV 4969
Qy 1290 SKEDGRSSGPPHETAAKRYDYMMEGRVGRAISSASTIEGLMGRAPPERHSPHILKEQH 1349
Db 4970 SREHFEPE--DHSHVRSQEY-----SAS-----GSPVPEK-----SVR 5004
Qy 1350 HIRGSIITOGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAVKQALGLPLKLP 1409
Db 5005 VIETTTTTVTRHEFDTEEDYIPSESRTHDDGITDQHVPSQS----- 5047
Qy 1410 AHEGLVATVKEAGRSIHPIREELRH--TPELPLAPRLKEGSIITOGTPLKYDTGASTTGS 1468
Db 5048 -----FVPEEDDHVTHQIIRDDPIDE--QIVESHKEYKSESSIPSPSQD 5089
Qy 1469 KKHVRSUIGSP--GRTPFPVHPLDMADA-----RALERACYEESL-----K 1509
Db 5090 STHVLETVMSTPITSERYDPEKDVESADDEIDSTAQYKSESPVQTEKSLILAKQ 5149
Qy 1510 SRPGTAS-----SSGSGIARG--APVIVPELG-----KPRQSLTYED 1545
Db 5150 QESGDESDEGFGSKVLGFAKKAAGVAGVVAAPVALAAVGAAYDALKKDDDEEDQED 5209
Qy 1546 HGAPAGHPLRGSPVTRMEPT---PRLQE--GSLSSSKASQKRLTST-----PREI 1592
Db 5210 KEPLIGGFHKQDDPISQDEPNAVDETPGATFVEPEEDKVIITDSADSSVQDEPKIV 5269
Qy 1593 AKSPHSTVPEHHHPHPISPYHLLRGVGVLDYRSHIPLAFDPTSPRGIPLDAAAAYLP 1652
Db 5270 --FPVDSTPEHNSNDRESFESIVKSEG----- 5294

Qy 1653 RHLAPNTPHYPPYILRG--YPTAALENROTIINDYITSQQMHNHTATAMA--QRADM 1709
Db 5295 -----PYIVESTDYAQTSAEER-----ISSPVHSDAGDASSFKRPDES 5332
Qy 1710 LRGLSPRESSIALNYAAGPRGIIILSQVPHLPVLVPPPTPGTPTATAMDRLAVLPTAPQPS 1769
Db 5333 VTGDEKNA-----IPETSETDAPV-----IDSEEFBN 5362
Qy 1770 SRHSSPLSPGPGTHLTKPTTSSSERDRDRRE-----RDRDRREKSLITSTT 1819
Db 5363 NRDEQRISPP-----AHSDEDEDAVIDSEFYRHSQEQNNEEDPSIV----- 5406
Qy 1820 TVEHAPIWRPTEQSSSGSSSGSGSGSSSPASHS-----HAHQHSPISPR---TQDA 1870
Db 5407 -----ESGEYISSGHG---SPRPPEDSITTTVLNVHHEPAALPEPEVDEE 5449
Qy 1871 LQQRPSVLNHTGK-----GIITAVEPKPTVLRTSTSSPVPRPAATFPFATHCPLGG 1923
Db 5450 LEQERSIIESEYKTSPLPPTSVTVEHVPEABIKYRTTSTP--IVTTVSSSEH----- 5502
Qy 1924 TLD--GVYPTLMPEVLLPKAPRVARPERPRADTG-----HAPLAKP--PAR 1966
Db 5503 -LDRNQEPYVVESEYTRASP--LGPERSPSPSPLPREDDSHVIESHYTSSPVSE 5559
Qy 1967 SGL-----BPASSPSKSGSPRPLVPPVSGHATIARTPAKNLAPH--HASPD-- 2010
Db 5560 DSVKHVIEKTTTTVTVEERPEDESHSPVSEDDV--HGFVKTTTTTTTTVTHEHFEPEDH 5617
Qy 2011 -----PPAPPASADPHREKTQSPFISQIQLSLRSLGVH---GSSYSPEG 2052
Db 5618 TSDEHVVESERYASGSPVPSENSNRVTETTTTTVTREHFEPEDDQDHVVESQEYSAG 5677
Qy 2053 VEPVSPVSPSLTHDKGLPKHLE-----ELDKSHLEGELRPKQPGVKLGGEAAHLPLR 2107
Db 5678 -----SPVPS-----EKSVKVIETTTTTVTREHFEHE--DEITIV-----ESSH 5717
Qy 2108 PLPESQSSSPLLQATPGVKGHQRVVTIAQHISEVITQDYTRHH---PQQLSAPLAPLY 2164
Db 5718 --DDQAASS--VPSEEDVHQIQT-----TTTTVTRHVVPDEIDSGRMDELE 5764
Qy 2165 SFGASCPVLDLRPPS-----DLYLPPPDHGAAPARGSPHSEGGKRSKSP 2207
Db 5765 KYSSSES-PV-----PPEDSSRVETTTTTFIREHFEPEDDHSHVVGVSQEYSAG--SP 5816
Qy 2208 EPNKTSV 2214
Db 5817 VPSEKSV 5823
RESULT 83
O60382 PRELIMINARY; PRT; 1791 AA.
ID O60382
AC O60382;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein KIAA0324 (Fragment).
GN KIAA0324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA Ueng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RT "Sequencing of Human Chromosome 16p13.3";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RESULT 84
Q9W320 PRELIMINARY; PRT; 3166 AA.
ID Q9W320;
AC Q9W320;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG3950 protein.
GN CG3950.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey L.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimball B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodge T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhou Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000).
RL [2]
RN SEQUENCE FROM N.A.
RP Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banjon J., An H., Baldwin D., Benson J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frick E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragaa V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";

Db 862 AGGSVTTTTTTITKSSPAPAPSP---TTAAPTSPASSAPPDNKLSQVYTTTTKPGD 918
Qy 807 ---TPPPAPSPSPAPPVVPKKEEETAAAPPVEEGEQKP---AAEELAVDTGKABEP 861
Db 919 IFSLPPTPTPTINNEPTLTTRNTTTTTTTTTTTATSDNLQNHKPSAIEPATDTD--SQP 976
Qy 862 VK---SECTEAEAGPA-----KGDAAEAATA-----EG 889
Db 977 IRKVKLSANEAKVVEAEFCVRQYYQLNGENPETPESCTPANKPKHOMRRPHDEPEP 1036
Qy 890 ALKAEKEGG---SGRATTAKSGAPQDSDSATSAD---VDEAEGDKNRLLSPR 941
Db 1037 QLRSSKSPSVEPQVQREITTFEGRRVSQDRE---ISIDELILIEETSGAPG---SPK 1088
Qy 942 -PSLTLTGDPRANASPOKPLDLKOLKORAAAIPPIQVTKVHBPREDAAPTKAPAPP 1000
Db 1089 IPS-----PRAQ-SPGKPATRSQ-----SPEKQPRATP 1116
Qy 1001 PPONLOPESDAPQPSGPRKSRSPAPPADKEAFAEAOGLPGDPPCWTSGLPFPVPPR 1060
Db 1117 ---ROSPEQOFAFKTHEVYTPSQEKFPRARSPEKTPG---WTO--FOVSPRQ 1164
Qy 1061 EVIKASHPAPPSAFSAPPGLPLGLHDHTARVLPRLPPTISNPP--PLISSAKHPSVL 1118
Db 1165 SPEKQLPRAQSEK---VPAVRQPSVSQPSPEKQIPDKTRDQOGPLPLISPRQSP--- 1218
Qy 1119 ERQI-----GAI SQMSVOLHVPVSEHAKA 1143
Db 1219 EKQLPKDVPQKSRQSPKDLTNQRRBEEIFRSTTTTTTKRTTNLNEEFITNERDNQ 1278
Qy 1144 PVGPVTMGLPLPMDPKKL-----APSGVQEOQLSPRQAGPESLGVPTAQASVLRG 1197
Db 1279 PISEKKQIPANAENTPKSETIESPDGPFPSKTTEVEAQ---PEVKESPYRKGLR 1335
Qy 1198 TA-----LGSVPGSGITKIPSTRVPSDSAITRGSITHGTPADVLYK----- 1240
Db 1336 ETFFEDRCQILGMEEDGD--TQTYTERPNNEQEDV---NVSHTTIETIQVKIEDCPNDDE 1391
Qy 1241 -----GTITRII-----GED-----SPSLDRGRED--LPKG--- 1266
Db 1392 DDKRRVTYTVVTRTPQKIKVBEELFVDVTEADEIVLNFPSKSPKEEDSPKYPKGPET 1451
Qy 1267 -----HVIYEGKKG-----HVLSEGGMSVTQCSKEGDR----- 1295
Db 1452 PKSPNDORIPSPKKGSPVOFKTEETPRYPQOEERYPKPETSQYKESPRNPKEADAE 1511
Qy 1296 -----SSSGPPHETAAPKRTYDMMEGRVGRA-----ISSASIEGLMGR 1333
Db 1512 TININEETIVITKEGSKSPSPWSPSPER-----RVPKSQPPSPASPSVPSVGR 1564
Qy 1334 AIPPERHSHLHKBQHHIRGSIQGI PRSYVEAQEDYLREAKLLKRGSTPPPPPSDL 1393
Db 1565 KIPNEVESNFVTEKIIDCRG-----KTWVB-----KISORPTSPPTPKNT 1607
Qy 1394 TEAYKTQALGP--LKLPAHEGLVATVKEAGRIHBIPEELRHTPELPLAPRPLKE----- 1448
Db 1608 KFSQKIPERVPETESEPEKDSSETKTTSVSVTKETERNSRTTKTKQPLPKEPOSK 1667
Qy 1449 -----GSITQCTPLKYDVTGASTTGSKKHDSVSLIGSPG----- 1481
Db 1668 VPAGSKPRKSLTGRKRDLSVBETRITTTTTTTQGRKPSDTN---GSPSFKDLRLSSPR 1724
Qy 1482 -----RTFPVHPL-----DVMDADALERACYEESLKRPGTASSGGSIGARGAPV 1528
Db 1725 KOKTSPQOQRTTPAQTRNPDVDGDSGSS-----PDASPTRVGNERRRNSNIVHTEI 1778
Qy 1529 IVPGLKPRQSLPTLVDHDCAPFAGHLP---RGSPVTMRE---PTRELQEGSLSSSKASQD 1582
Db 1779 IDHMAP--KSPKT-ERRSQGTGVVPSPIRKLPLTERKESAPVRV-----TERDKA 1828
Qy 1583 RKLTL-STPREIAK-----SPHSTVPEHHPHIPISPYEHLRLGVSGVDLYRSHIPLADPTSI 1637

Db 1829 EKVTRSTSENIIKMSGTHKPHPEMSLKGPGDRSRPSKCTTKTINLSEORINTATD----- 1884
Qy 1638 PRGIPLDAAAAYVLPRLHAPNPTYPHLYPYLIRG---YPTDAAA----- 1678
Db 1885 IEGVIIIDIOQA-KSSREPSDRIVTPVPAELETGKPRIPDVVOEPDDDEPRKRPQVTNIP 1943
Qy 1679 --LENRQTTIINDYITSQOMH-----HNT 1699
Db 1944 IFEEESQTYVGCQIS--ELHSSNGIEVDILDNPTVEAPKSLDYVNVPTPDTDESLLSVHEK 2001
Qy 1700 ATAMAGRAMLRLGSLPRESS-----LALN-----YAAGRGLIDLS 1735
Db 2002 VSRFTSHABKVK--EPKVSAPFSREDFVNAKIPENDCLLSINOKYDKFLRTAENVIRPT 2059
Qy 1736 QVPHLPVIVPPTGTP-----ATAMDLAYLPTAPOFPSSRHS 1773
Db 2060 SLPSRPEI--ERPELBEIDELRDCTLSVSKVHKFIDTAEKLA--PTMFO-----K 2109
Qy 1774 SSPLSPGGPHTLTKPTTTSSSERERDRDRDR-----BREKSILSTTT 1820
Db 2110 SPRIVANIERHISR--QSEPERELDESEPELDRDVEDDDQTSQLETEEBEITQTVTK 2166
Qy 1821 VEHAPIWRPOTEGSSSGSSSGSSSGSSSRPASHAHQSPISPRTOALQORPSVLHN 1980
Db 2167 KETLKEFKQOTKETRETRRDSKAEPEKLOKSPQTKVKEESARVPKYQAKVOK----- 2220
Qy 1881 TGMKGIIITAVEPSK-----PTVLIRSTSTSPVRPAATFPFATHCPLGGLDGVYPTLMEP 1935
Db 2221 -----VSQWEPKKQOREPKV--TQKETPLFPKKQ-----PLSKVKD-----EP 2257
Qy 1936 VLLPKAPRVAR-----PERPRADTGHAFILAKPPAR--SGLEBPASSPSKGSBPRL 1984
Db 2258 EKVKREPKVPQESQTKLEEPEVTKTPOKEPRKEPLRQSEDEPEFSPSEEFDDDEPL 2317
Qy 1985 VPPVSGHAT-----IARTPA--KNLAPHASDPDPAPPASASDPHREKTSQSPFSI 2033
Db 2318 -PMTKHTHTTAEMKROKDILNRPSVFGORTPEKSSSTPSP-----TKLNGTRGRSPS 2370
Qy 2034 QEL---ELRSLGVHSGSYSPGVEPVSP--VSSPSLT---HDKGLPKHLELDK-----SH 2081
Db 2371 TNLITEKSYRQVNTVNSKPGTKTTPSANSQAQSPPTTTSIKRMEQISQGSVWVQD 2430
Qy 2082 LEGELRPKQPGVKLGEEAAHPLHPLRPESQ---PSSSPLLOTAPGVKHQHVTLAQH 2138
Db 2431 VDQDVEVVGAPPSP-----HISEKPOKSPSTSSRSLSRSPSRSPSRETSTNLNT 2481
Qy 2139 ISEVITODYTHHPQQLSAPLAPALYSPGASCPVLDRPPSDLYLPPDHGAPARGSP 2198
Db 2482 TSTNTTT--TTEHPSTTKTTTPKPTSTNPKDBPEI-----IP 2517
Qy 2199 HSEGGKRSPEPNKTSVLGG-----GEDGIBFVSPPEGMPTEP-----GHRSASVYPLL 2245
Db 2518 IESLTKESIITTYTNTTGRNVASRRNVFPVETHVDSEPTGRRRPSYMDHTKSSL-EHI 2576
Qy 2246 YRDEQTEPGRMSGSPGNTSQPPAPFSKLTESANSAMVSKKQKINKLKNHNEP--- 2302
Db 2577 RRDLSLEINKSHYRSKSMEDDSP-----VEPRNPNS-----SVKFDVPRKSSRGDAEPKXT 2628
Qy 2303 -----EYNISOPGTIEFNMPAITGGLMTYRSQAOE-----HASTNMG 2341
Db 2629 SLKGKSDSDLEIEIEIFDLQRLKLE--LETVASIEMRRIRRAQMRILKMNINAGTTTT 2687
Qy 2342 LEAIIRKALMGKYDQWEEESPPLSANAFNPLNASASLPAAMPITAADGRSDH-----TL 2394
Db 2688 ITTITTTSTPCK-----SSPLPKIRRDQSPAGAAEVKTEVTTTTSRQOQOQVQVDS 2741
Qy 2395 TSPGGGKAKVSGRPSR---KAKSPA-----PGLAS-GDRPPSVSSVHSEGDNR-- 2441
Db 2742 TTPIAPGKTSPHGKPPVKPRERSASPAQKRIRISPGKQSPGDRSTTTTTTKVTITSTTRGA 2801
Qy 2442 -RPLTNRVNEDR-----PSSAGSTP 2461
Db 2802 PSKPAQGPFWADRSKVLKGHATVPQTNGSTP 2832

Db	556	P\$ASPQERSESDSPD\$KPKTRTPLRQRSHSGSSPEVD\$KSRH\$PRLSRSGSSPEMDKPK	615
Qy	722	EALHASGNEVREGCSGPATVNN\$SDTESISP-----HTEAAKDTGQNGPKPATLGCAD	776
Db	616	RVLQRAQ-----SG-----TDSSPEHKIPAPRALPHRSRSGSSSKERGPSPGSSSSE	663
Qy	777	GPPGPP--TPPRRTSRAPIEPTPASEATGAPTPPPAPPSAP\$PVPVVK\$EKEETAAA	834
Db	664	SSPHAPK\$RTARRGSRSSIEP-----KTK\$RTPPRRSSRSP-----	702
Qy	835	PPVEGEQKPPAAELAVDTGKAEEPVKSECTEAE\$EGPAKGDAAEAATA\$GALKAE	894
Db	703	-----ELTKA-----RVS	711
Qy	895	KKEGSGRATTK\$SGAQDSD\$SATCSAD\$VD\$AEGDKNR--LL\$PRPSLLTTPG\$PRA	953
Db	712	R\$RSASSPEIR\$RTPPRRR\$FVS\$PEPTEK\$SR\$RRR\$SVSSPR-----TKT\$SRG	767
Qy	954	NASQKPLDLKOLKORAA\$PIPIQVTKVHEPREDAA\$TPKAP\$PAPPPQNLQ\$PESDAPQ	1013
Db	768	R\$SPKPRGLQ\$RSR\$R-----R\$KTRTR-----RRDRGSS	801
Qy	1014	QPGSSPCK\$RSPA-----PPADKEAFAAEQAQLPGDPPCWT\$GLPFPV	1058
Db	802	Q\$TRRRQR\$RSR\$RVTRRRRG\$GYSR\$SPTRQ\$ESR\$TSRRRG-----R\$KTL\$TSR	856
Qy	1059	PREVIKASPHADP\$AF\$YAP\$GHPPLGLGLHDTAR\$PVLPRPPTISNPPPLISSAKH\$SVL	1118
Db	857	K\$RSR\$TP--APWK\$RSR\$RSPA-----THRSR\$RTPL\$SR--\$SR\$TSFVS	903
Qy	1119	ERQIGAI\$QGM\$VOLHVPYS--EHA\$KAPGVPTMGLPLPMDPKLAP\$F\$GVKQ\$BQL\$SRGQ	1177
Db	904	RRRS\$VNR\$R\$SR\$R\$AG\$FVS\$RRR\$SR\$TPPVTRRR\$SR\$R\$TPTRRR\$SR\$T-----	953
Qy	1178	AGPESLGVPTAQ\$ASVLRTAL\$C\$VPG\$SITKIGIP\$TRVPSD\$AIYR\$G\$ITHGT\$PADV	1237
Db	954	--PP-----VTRRR\$SR\$TP--\$VTRRR\$SR\$TSFV--\$980	
Qy	1238	LYKGTITRIIGED\$F\$RLDR\$C\$R\$D\$LPKGHVIYEGKKGHVLSYEGGMSV\$TQ\$CKED\$RSS	1297
Db	981	-----TRRR\$SR\$TP\$VTRRR\$SR\$TSP-----VTRRR\$SR	1012
Qy	1298	SGPPHETAAPKRTYDMMEGRVGR\$AISASIEGLMGR\$AIPP--ERH\$PHILKEQH\$HIRGSI	1355
Db	1013	T\$PVT\$RRR\$SR\$T-----PPAIR\$RS-----R\$R	1036
Qy	1356	TQ\$IPR\$YVEAQDY--LRREAKLLKREGT\$PPPP\$P\$RDLTEAYKTQALGPLK\$PAHEGL	1414
Db	1037	TPL\$PRK\$RSR\$PLAIR\$RS-----R\$RT\$PRAAR\$K\$R\$SLTRS-----PPAIRR	1081
Qy	1415	VATVKEAGRSIHEP\$RELRH---TPELP-----APRLKE-----	1448
Db	1082	SAG\$G\$DR\$RS\$ATPPATRNH\$G\$RT\$PVALSSR\$M\$CFR\$P\$M\$PT\$P\$D\$CR\$P\$G\$M\$LEP	1141
Qy	1449	--GSITQGTPLKYDTGASTG\$K\$KH\$DVR\$SLG\$SPRT\$PPPH\$PLD\$VMA\$DARALER--ACY	1504
Db	1142	LGSART\$M\$V\$LQ\$TGG-----SMDG\$CPRI\$PD--HPR\$SVVENH\$AQ\$RIALALT	1189
Qy	1505	EESL--K\$RPGT\$ASSG\$GIARG\$PIV\$P--ELGK\$PQ\$P\$LYTEDH\$GAPFAGH\$LP\$G\$PVTM	1562
Db	1190	AISLGTAR\$PPPSM\$AAGLAAR\$MSQVAPVPL\$M\$SLRTAP-----AANL\$ASRI\$PAASAAAM	1243
Qy	1563	REPT\$PRLQ\$GL\$SS\$K\$AQ\$KLT\$T\$PRETA---K\$PH\$STV\$PEH\$PH\$P\$IS\$PVEHLLRGV	1618
Db	1244	-----NLASART\$AIPASVNLAD\$RT\$PAAAAAMNL\$ASPR\$TAVA---PSAVNLAD\$PRT\$PAA	1295
Qy	1619	SGVDLYR\$HIP\$AD\$PT\$IP--RGIFL\$DAA\$AYL\$PRH\$APNTP\$VPHLY\$PPY\$LY\$RG\$PDTA	1677
Db	1296	SAVNLAGART\$PAA\$LAAL\$SLTG\$TGP\$TAA-----NYP\$SS	1330
Qy	1678	ALENRQTIINDYIT\$SQ\$MHN\$TATAMAQRAD\$MLRGL\$SPR\$SS\$ALN\$YAA\$GR\$GIIDLSQV	1737
Db	1331	RTPOA\$PT\$PAN--LVVGR\$P\$A\$G\$TAPVNIAG\$RT\$PAGL\$AP\$TNL\$SS-----RMA	1376

RESULT 87

RESUL 67
095KU4

ID 095K

AC Q95K

DT 01-D

DT 01-D

DT 01-0

DE Trea

	GN	TCOF
1	0.0000	0.0000
2	0.0000	0.0000
3	0.0000	0.0000
4	0.0000	0.0000
5	0.0000	0.0000
6	0.0000	0.0000
7	0.0000	0.0000
8	0.0000	0.0000
9	0.0000	0.0000
10	0.0000	0.0000
11	0.0000	0.0000
12	0.0000	0.0000
13	0.0000	0.0000
14	0.0000	0.0000
15	0.0000	0.0000
16	0.0000	0.0000
17	0.0000	0.0000
18	0.0000	0.0000
19	0.0000	0.0000
20	0.0000	0.0000
21	0.0000	0.0000
22	0.0000	0.0000
23	0.0000	0.0000
24	0.0000	0.0000
25	0.0000	0.0000
26	0.0000	0.0000
27	0.0000	0.0000
28	0.0000	0.0000
29	0.0000	0.0000
30	0.0000	0.0000
31	0.0000	0.0000
32	0.0000	0.0000
33	0.0000	0.0000
34	0.0000	0.0000
35	0.0000	0.0000
36	0.0000	0.0000
37	0.0000	0.0000
38	0.0000	0.0000
39	0.0000	0.0000
40	0.0000	0.0000
41	0.0000	0.0000
42	0.0000	0.0000
43	0.0000	0.0000
44	0.0000	0.0000
45	0.0000	0.0000
46	0.0000	0.0000
47	0.0000	0.0000
48	0.0000	0.0000
49	0.0000	0.0000
50	0.0000	0.0000
51	0.0000	0.0000
52	0.0000	0.0000
53	0.0000	0.0000
54	0.0000	0.0000
55	0.0000	0.0000
56	0.0000	0.0000
57	0.0000	0.0000
58	0.0000	0.0000
59	0.0000	0.0000
60	0.0000	0.0000
61	0.0000	0.0000
62	0.0000	0.0000
63	0.0000	0.0000
64	0.0000	0.0000
65	0.0000	0.0000
66	0.0000	0.0000
67	0.0000	0.0000
68	0.0000	0.0000
69	0.0000	0.0000
70	0.0000	0.0000
71	0.0000	0.0000
72	0.0000	0.0000
73	0.0000	0.0000
74	0.0000	0.0000
75	0.0000	0.0000
76	0.0000	0.0000
77	0.0000	0.0000
78	0.0000	0.0000
79	0.0000	0.0000
80	0.0000	0.0000
81	0.0000	0.0000
82	0.0000	0.0000
83	0.0000	0.0000
84	0.0000	0.0000
85	0.0000	0.0000
86	0.0000	0.0000
87	0.0000	0.0000
88	0.0000	0.0000
89	0.0000	0.0000
90	0.0000	0.0000
91	0.0000	0.0000
92	0.0000	0.0000
93	0.0000	0.0000
94	0.0000	0.0000
95	0.0000	0.0000
96	0.0000	0.0000
97	0.0000	0.0000
98	0.0000	0.0000
99	0.0000	0.0000
100	0.0000	0.0000

OS Cani

OC Euka

OC Mamm

OX NCBI

RN	[1]
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

RP	SEQU
----	------

RA Hawo

RA Edwa

RT
pm

RT	dogs
pr	comb

RL	Subm
DP	EMPT

DR	EMBL
DR	Total

DR Inte
DR Inte

DR	Inte
DR	Pfam

DR PLAIN
DR PRINT

DR. PRINCE SMAR

DR	EMFAR
DR	PROS

SO SEOU
FRON

30

Query M

Best Lo

Matches

Qy

Db

1

Db 106 RLASTNSVPGVLPVPSSTKEKGVAKTNKAKKQVNS--TTHPASAKAVAHILSGSPRKSAG 164
Qy 819 PP--VVPKEEBE-----TAAAPV-----BEGEOKPAAAEELAVDTGKAEEPVKS 864
Db 165 PSANTILVSETEEGSVPALGTTAKPGMASANQADSSSETSSSDDETDVEKASEKIVQA 224
Qy 865 ECTEAEAGPKAGKDAEAAEATGAALKAEEKGGSGRATTAKSGAPQDSDSATCSAD 924
Db 225 ---KAASGPVKGTGKATPAPP-----KAGFS--AAQAKTEKPKEDSDSS---EED 269
Qy 925 EVDENAGDKNRLLSPRSLTTP---TGDPRANASPKQLDLKQLKQBAAIPIQVTKV 981
Db 270 SDSSEPPAAKTPLOVKPSGKTPQVKAASAKESPRK-----GVPPVPPGKV 317
Qy 982 -----HEPPREDAAPTAP-----PAPPPQNLQ-----PESD 1010
Db 318 GPAQAKKAGGEDDDSTESSEEAFTAVPTTRSPVQAKFSGQNSQVRTASGPVK 377
Qy 1011 APQO--PGSGPRGKSRSPAPPADKEAFAAEAKLPGDPDPCWTSGLPFPVPPREVIKASPH 1068
Db 378 PPQKAGPAATPVGKEE-----DSESSSEESDSEGAAPAQA-----KSSGK 419
Qy 1069 APDSAFYAPPGHPLPLGLHDTAPVLPPTTNSNPPPLISSAKHPSVLEROIGALSOG 1128
Db 420 IPQVRAASGPAGK-----PPQKAGPAATQAKAK-----MSKDDSESEEE 458
Qy 1129 MSVQLHVPVSHAKAPGVPTMGLPLPMDPKLAPFSGVKQEOQLSPRCQAGPPESLGVPT 1188
Db 459 ESES-----ESEAAMAPVQAKSTVTPQTKASPKGT---PTTPPAKAPPVVRGTGA 510
Qy 1189 AQEASVLRTALGSPVGSITKGPSTPRVPSDSAITYRGSITHTGTPADVLYKGTITRIIG 1248
Db 511 PWKARAEASPCASSP--AMARGAQKPEASSS-----541
Qy 1249 EDSRRLDRGREDSLP---KGHIYEGKKGHVLSYEGGMSVTQCSKEDEGRSS-----GP 1300
Db 542 EESES-----BEETAPAPAAGQKPVGK-----GVTYKAASPTPTKPSGOGTALAP 587
Qy 1301 PHE---TAAPKTYDMMGRVGRVIAISSASIEGLMGRAITPPE---RHSPHLKQHHIRG 1353
Db 588 POKAGPATAPVKTVEQEDSSSE 647
Qy 1354 SITOG---IPRSYBAQEDYLREAKLLKREGTPPPPPPSR-DLTEAYKTOALGPKLKP 1409
Db 648 TSAKGAASAPGVGSA-----AVQIKQESPAKVKKPPARTSQSNVSVRGQGA---P 696
Qy 1410 AHGULVATVKGAGRSIHEIPRELBHTPELPLAPRLKEGSIQTQTPLYKDTGASTGSK 1469
Db 697 AVGKAVATAAQA-----QPGPVKNPEEDSESESE 725
Qy 1470 KHDVRESLIGSPGRTPPPVHPLDVMADARALERACVEESLKSRRPGTASSSGGSIARGAPVI 1529
Db 726 ESD-----SDGEATPVKP-----SGKTQVGTASAPSGLGRKGATA 763
Qy 1530 VPGLKPKRQSPLYTHDGHGAPFAGHLPGRSPVMTREPTPRLOEGSLSSSKASQDRKLTSTP 1589
Db 764 AP---PKTGAVATAQAKP-----EEDSESESESESDSE-----794
Qy 1590 REIAKSPHSTVPEHHPHISP-----YEHLLRGVSGVDLYRSHIPLAFDPTSIPIRGIP 1642
Db 795 -----EEAPQVKPGIKPQTKASLLKGVSGT-----P 822
Qy 1643 LDAAAYVLPRLAPNPTVPHLYPPYLIRGVPDTAALENRQTIINDYITTSQMHNTATA 1702
Db 823 ASAEASTTRVDNSAPQKARP--APP-----AKEGSKTKAKS 857
Qy 1703 MAQRADMLRGLSPRESSIALNYAAGRGHIDLSQVPHLPVLVPP-----TPGTPTATA 1754
Db 858 KAQAP-----APPEKN-----AEGSSESSD-DELPATQVIKPPPLIFVDPNKRSPAGPAAT 905
Qy 1755 MDRLAYLTAPOPFSSRSSSSLSPGCGTHLTKPTTSSSERDRDRDRDRDREREKSI 1814
Db 906 ATQ---APAAADTPRKAQASES-----TARSSSESEDED-----V 937

Qy 1815 L-----TSTTTVEHAPIWRPCTEQSS--GSSGSGGGGSGSSRRPASHAHQHSP 1862
Db 938 IPATQCSVPAGRTNVITPLTAHP--RPAVRASTVGASGGEASGRVSEK-----KQEAP 988
Qy 1863 ISPTQDALQOORSVLHNTGKGIITAVEFSKPTIVLRSTSTSPVRPAATFPATHCPLG 1922
Db 989 ITQVTKENPAHLP--LTQAAKLVLAQKASEAQPPAARTPSSSGVDHALGTLF-----1038
Qy 1923 GTLDGVYPTLMEPVLLPKPEAP--RVARPERPRAD-----TGHAFKAPKPARSGLBPAS 1973
Db 1039 -----VMSQITPVQAKMTNKLKAEVPAVERATATPVGHPKAKASETSDNSEDSS 1089
Qy 1974 SPKSGSEF-----RPLVPPVPSGHATTAR-----TPAKNLAHHASPDPP 2012
Db 1090 GSSSGSEEDADGPKAHRVGPAPSATETILVEETABESSEDEVVAPQSLLSGVVTGPT 1149
Qy 2013 APASADPHREKTQSKPFSIOELESRLGYHSSYSYSGVEPVSPVSSSLTHDKGLPK 2072
Db 1150 LANSQAS-----KATPRP-----DLNP--SASSTSIKDAPD 1179
Qy 2073 HLEELDKSHLEGELRPQPGPKVLGGEAHLPHLRPLPESQSSPLLQTAPGVKGQHV 2132
Db 1180 GKQEVPEQHVAGTVSPKTS-----RREADATPQ--KPRKPKKEAGSP-----QA 1221
Qy 2133 VTLAQHISEVITQDTRHHPQOLSAPLAPLYSPFGASCPLV-----DLRRPPSDLYLPP 2187
Db 1222 STLA-----LQSDISR---RLISEPWPLNEAQVQASVWKVLTLEBQERKKAADAAKES 1272
Qy 2188 PDHGAAPARGSPHS--EGGKSRSPNKTS--VLGGEDGIEFVSPPEGWTETP--GHSRSVY 2242
Db 1273 NRKGRVGRKRLSGDQTAARVPKSKKKQLVAGGGGGAAG--SPEKALRTPKGAK-----1327
Qy 2243 PLLYRDGEOTE--PSRMSKSPGN-----TSOPPAFFSKLTESNSAMVKSQKQEIKNKL 2294
Db 1328 ----RDGASDIEKKKEKESFSGLEAKEKEGEPGTCLKVGGDQGNFKIKEKKSKDKK 1383
Qy 2295 NTHRNEPEYNIQPGTE 2312
Db 1384 KDEKKEKKKKKAKKASTK 1401
RESULT 88
Q966V1
ID Q966V1 PRELIMINARY; PRT; 2376 AA.
AC Q966V1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE REXIN L1.
GN CAP OR RXN OR CG3451 OR CG18408 OR CG18409.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamazaki H.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yamazaki H., Willert K., Fish M., Nusse R.;
RT "Drosophila REXIN, a Novel SH3 Adaptor Protein of Axin and Arrow that
is Essential for Living in Late Stage Embryo.",
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 3 SH3 DOMAINS.
DR EMBL; AB053478; BAB62017.1; -;
DR Flybase; FBGN0033504; CAP.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF000018; SH3; 3.
DR PRINTS; PR00499; P67PHOX.

DR ProDom: PD000066; SH3; 3.
DR SMART; SM00326; SH3; 3.
DR PROSITE; PS50002; SH3; 3.
SQ SH3 domain.
KW SEQUENCE 2376 AA; 267657 MW; 9C7BD6C7A705C888 CRC64;

Query Match 3.1%; Score 409; DB 5; Length 2376;
Best Local Similarity 18.9%; Pred. No. 4.3e-11;
Matches 481; Conservative 273; Mismatches 826; Indels 960; Gaps 114;

QY 60 QPQRRPRLSEFQPGNRSQSE---LHLRPESHSYLPGLGKSE-MEFIESKRPRLELLP 114
DB 20 QNQONPS-----QNQNNQOQAEGADREBQQDFQIAVAQSSFVDN----- 66
QY 115 DPLRPSLLATQGPAGSEDLTKRSLTKGLEPVPSPPPHTDPELELVPRLSKEELIQ 174
DB 67 ---GNSGSVNGPTENSEEVNTLEKTEKEEIF--EQPAT-----VIQ 105
QY 175 NMDVRDREITWVEQIQSKLKKQ---OLBEEAAKPP----- 208
DB 106 KEADSDAEMGAKNSKHKREKSKDQASNGKSEBQVRRPPTIIRPPGSPQAKVIVHRTVR 165
QY 209 -----EPKXPVSPPTIESKHSLSVQIYDENRKKAEAAHRLLEGLGPOVELPLYNOPS 261
DB 166 BEDTANGRAQSPPEPRAVESKEQHME--VEDSKKEKPEAQOEI-----ETKEREKEPS 217
QY 262 -----DTRQYHENIKINQAMRKKLILYFKRRNHARKQWKQFCQRYDQLMAL 309
DB 218 PKELPESHLSEDAQKHVVEHQ-----QVYD---EVD 249
QY 310 EKKVERIENPRRAKESKVEYY---EK--QFPE-----IRKORELQERMQ 351
DB 250 YSKDETTONNQOEQHSKPEQYVLOLEKAMEFVENAHQHQHVAQVQYQKQORDTEQNAQ 309
QY 352 SRVGQSGLSAARSSEHSE---IIDGLSEQENLEKQKQLAVIPMLYDADQ--- 405
DB 310 QOEONEAHLQESTKAHQONIAVPOILVDPKKREGLAQHHDVKEEPIILNNAQQQND 369
QY 406 -----RIKFNNGLMADP-----MKVYKDVQVMMSQEQ--KETPRE--- 442
DB 370 TAPKSPKEVTIKVQFOTEDPVSQPPQPKPTSKVLIHQHVETTDQNAKPTVEVSST 429
QY 443 -----KPMQHPKPF--GLIASFLERKTVACVLYYYITKKNYKSLVRSY 487
DB 430 TGSAGLTSPRPYLVESPKNVSSQFNSFRNVQIQEWEL-----NSDFSSEFNSL 482
QY 488 RR-----RKSGSQOQQOQQOQQOQQOQMPRSSQEEKDEKEKEAEKEEKEPVENDK 543
DB 483 QSLVCEVDSSEGSVILGPERSPSDQVPPSSSRVQEHEQVRQKRA-----QYRNAL 535
QY 544 -----DLKKEKTDGSDNDEKEAVASKGRKTANSQGRKGRITRSMANEANGSEAITP 598
DB 536 SHFMPQLLNPRYLDLSILEENWRNSTASG--GSDQTGIR-----TP 575
QY 599 QQSALASMEINSGSRWTEEE-----METAKGLLEHGRNWSAIAARMVG--SKTVSQ 649
DB 576 KLNFTFPRSQDQFRRKRRREAPSPKLETK---LLEESTDLSECTRLQALSQSEDA 632
QY 650 KNFYPNYKRONLEILOQHKLMKEKERN-ARRKKKAPAAASEAAAPPPVVEDEMEAS 708
DB 633 ELVLSSASSVSDLME---LELQAAALALVLDLTDASRLIARP-----DDEMSSTS 686
QY 709 GVSNGEEMVREAEALHASGNEVPR--GCSGPATVNNSSDT-----E 749
DB 687 TTTADRSTEATE-----TETEREGQSRESTDVNNANPTLASSQSLLSAATPTPTPT 740
QY 750 SIPSPTBAKDTGNGKPPATLGADGPPFGPPTPPRPTS---RAPTEPTPASEATG-- 804
DB 741 PTPADREQLKDTNVSGS--TVSFGAASPLAATREFVRNMDKVRELIEMTRREQECEL 799
QY 805 ---ATPPPAPSPSPAPPPVVKKEKEEBETAAPPVVEGEBOKPPAAEBLAVDTGKAEPV 862
DB 800 PRSPSPPPVPPPPASVPSPSPES-----SSFHLASLQLKQBSN 838

QY 863 KSECTEAEAGPAKQDAEAAABATAEGALKAEBKGGSGRATTAKSGAPOQSDSATCS 922
DB 839 DSHCS-----DSTHSQCT 852
QY 923 ADEVDDEAGGDKNRLLSPRPSLLTPTGDPPRANASPKPLDLKQLKQRAAAAPPIQVTKVH 982
DB 853 A-----INLASPPP---PPTAQFPTPLRQKP-----APFPV----- 881
QY 983 EPPREDAAPTKPAAPPAPPPQ---NLQPEDA-----POQ 1014
DB 882 -----PQVSPAPPTPQSEPELSVADSVANANADVADADTDTAIKRLRLCTEQ 933
QY 1015 PGSSPRGKSRPAPPADKEAFAAEQAQLPGPPCWTSGLPFPVPVPREVIKASPHAPDPSA 1074
DB 934 LASMPYGBQ-----VLEELASVAQNISDQS---QNKMPYMPQLPHIKELQLANESK 983
QY 1075 PSYAPPHPLPLGLHDTARPVL-----PRPTISN-----PPPLISSAKHSPSVLERQIGAI 1125
DB 984 SSSA-----WGLPTQSDPKLLVCLSPGQDLVNNQTPDDLLDA--HOKFVERR----- 1031
QY 1126 SOGMSVOLHVPYSBHAKAPVGTWGLPLPMDPK-----KLAPFSGVKOBQLSPRQA 1178
DB 1032 -----GYHELUSKAQV-----LEQHQOQQOQSEMUKTAAAMRELKSLSPAPP 1073
QY 1179 GPPESLGVPTAQEAASVLRGTALGSPVPGSITKGIPSTRVPVSDSAITYRGSIHTGTPADVL 1238
DB 1074 VPPPPVPVKSATAA-----KATAHENAKRDDASEQKQKIACESSSL- 1116
QY 1239 YKGTITRIIGEDSPSR-LDRGREDSLPKGHVIEGKGHVLSYEGGMSVTQCSKEDGRSS 1297
DB 1117 -----ENKPRRAADLG-----AQOASQROSS 1138
QY 1298 SGPPHETAAPKRTYDMMEGRVGRALSSASIEGLMGRALPERRHSPHILKEQH----- 1349
DB 1139 S-----TTTSSHKATTETMSSDTAKPPTLDSEBELARMFP--QHKGDIFEBQKRFSNTEF 1193
QY 1350 --HIRGSITQIGIPRSYVEAQEDYLRE-----AKLLKREGTTP-----PPP 1388
DB 1194 PSHQVPSQTKRYSNIETSSYESKKWENGQVYVDVSTSSHEKKEQDPPKQOPVPPVPP 1253
QY 1389 PSRLDTE-----AYKTOALGPLKPAHEGLVATVKEAGRSIHEIPRELRHTPELPLAPR 1444
DB 1254 PIMSATKLNGTIFDGDVAPKNQSPSRE---SGSGSGNGTYEFPQAKAAADAFGEQR 1309
QY 1445 PLKSGSITQCTPLK-YD-----TGASTTGSKKHDVRSGLSGPRTFPPV-HP-LDYM- 1493
DB 1310 EQQNG-LDQDRVFKDFDKLSQOMHABLOSTREKREKSASMYDLSGFTTPASGHPLDELQ 1368
QY 1494 ---ADARALERACVYERLSKSRPGTASSGSGSIARGAPVIVP---ELGKPR--QSPLTYE 1544
DB 1369 QRRHAHQLELE-IERSAKSRQERMSS-----VPRQMEATPPTTHEIPIELE 1415
QY 1545 DHGAPFAGHLPRG---SPVTMRBPTPLQBSLSSSSKASQQRKLTSTPREIAKSPHSTV 1600
DB 1416 -----PSRRAESLCNLNEPPPR-----PHTTV 1438
QY 1601 PEHHPHI-----SPYEHLRGVGDLYRSHIPLAFDPTSIPTGIPLDAAAAVYLPRLH 1655
DB 1439 -GHYNHPVVDWGRYAN-----DLGYSE-----NIARPFAREVEICYO----- 1476
QY 1656 APNPTYTH-LYPPYILIRGYPDTA-----ALENRQT---IINDYITSQOMHNTA 1700
DB 1477 RQNRTHICIRAPRLSASTNDLSSSSQVSYDTFNAYGGRTHAPMLNQAQQOORPHYGSC 1536
QY 1701 TAMAQR-----ADMLRGLSP-----RESSLALNYAAGPRG 1730
DB 1537 YSMIERDNPRIYISTTSRRGVSPAPPVATPQQOQVPPPAYDROQRSSL-----PRE 1589
QY 1731 IID-----LSQVPHLPVLVPTTGTATANDRLAYLPTAQPFSSRSSSSPLSPGGPTH 1784
DB 1590 LHEOQLKYLKSELKFEVERLOQERRRLMEQWAPVLPAP--QRRESYPAA----- 1642

Db 1099 KRHYKSVRSRREKTSRPLRLILNVSNGD-----RVBECOLETHNRKMTFKFD 1150
Qy 1293 -DGRSSGPPHETAPKRTYDMEGRCVRAISSAIEGLMGRAIPPERHSHPHLKEQHHI 1351
Db 1151 LDGN-----PEIATIMVNDFI-----LATERESFVDQVREII-----EKADEMUSEDVSV 1198
Qy 1352 RGSITQIGIPRSYVEAQEDYLRRREAKLKREGTQPPPPSRDLTEAYKTOALGPLKLKPAH 1411
Db 1199 EPEGDQGL--ESLQKDDYGFSGQKLEGEFKQIPASSM-----PQIGIPT 1244
Qy 1412 EGLVATVKEAGRS--IHEIPRELRHTPELPLAPRLKBSITQGTPLKVDYDGTASTTGSK 1469
Db 1245 SSLQVQVHSAGRRFIVSPVSRLESKVFP-----SEIT-----DTVAASTAQS 1289
Qy 1470 -----KHDVRSI-----IGSPQRTFPPVPHPLDMADARALE 1500
Db 1290 PGMNLSHASLSLQAFASELRRRAQMTGPNTPNFSHTGTPFVVPVPP----- 1338
Qy 1501 RACYEESLKSRPGTASSGSGGIARGAP-----VIVPELCKPROSPL--TYEDHGAPFA 1551
Db 1339 ---FLSIAIGVPTTAATAPVATSPNDISTSVIOSEVTVPTBEGIAGVATSGVVS 1395
Qy 1552 GHLPRGSPVTWREPTPLQEGSLSSKASQDRKLTSTPREIA---KSPHSTVPEHHPHPI 1608
Db 1396 GGL-----PFPVSESPVLSSVVS---ITIPAVVVSISTTSPSLQVPTSTSEIV 1441
Qy 1609 SPYEHLLRGVGVLYRSHIPLADPTSI PRGIPLDAAAAYLPHLAPNPTPHLYPPY 1668
Db 1442 -----VSSITALYPS-----VTVSATSASAGGSTATPGK-----PPA 1473
Qy 1669 LI-----RGYPDTAALENRQ-----TIINDYITSQOMH 1696
Db 1474 VVSQQAAGSTTVGATLTSVSTTTFFPSTASQLSIQLSSSTSTPTLAETVVVSAISLQKTS 1533
Qy 1697 HNTATAMA-----QRADMLRGISLPR-----BSSLALNVAAGPRGIID 1733
Db 1534 HSSITGLAFSLAPSSSSPCAGVSSYISQGLHLPLVPSVIASTPILPQAAGTSTPL 1593
Qy 1734 LSQVPHLPVLVPPFGTGPATANDRLAYLPTAPQFSSHSSPLSPGCP--THLTKPTTS 1792
Db 1594 LPQVPSIPPLVQPVANVPAV---QOTLHSQOP-----ALLPNQPTHCP----- 1636
Qy 1793 SSERDRDRDRDREREKSILSTTTVEHAPITWRPGTESGSSGSSGGGSSSRPA 1852
Db 1637 -----EVDSDTQKAPGIDDIKLEB-----KLRSLSSESSSA----- 1671
Qy 1853 SHSHAHQSPISPRTDALQORPSVLHNTGMKGI--ITAVEPSKPTVLSTSTSPVRPAA 1911
Db 1672 -----QHASVSLTS-----LVISTVTPGIPTTAVAPSK--LLTSTT-----S 1708
Qy 1912 TFPATHCPLGGTLDGVVPTLMEPVLLPKEAPVARPERPADTGHAFKAPPARSGLEP 1971
Db 1709 TCLPPTNLPLG-----TVALPVTVPVTCQVSTPVSITTSVGKVP 1747
Qy 1972 ASSPSKSEPRPLVPVSGHATARTPAKNIAPHASDPDPAAPPASADPHREKTQS--KP 2030
Db 1748 GTAFSKPPLTKAPVLVGTETLPAGTLPSEQ-----PFPQPSL-----TQSQOP 1792
Qy 2031 FSIQELRLSYGHGSSYPGVEFPVSPVSPFSLTHDKGLPKHLELDKSHLEGELRPKQ 2090
Db 1793 LEDLDAQLR-----RTLSPRIITVTSVAV----- 1815
Qy 2091 PGPVKLGGAHLPHLRPLPE---SQPSSSPLLQTAQCV---KHQORVVTLAQHISEVI 2143
Db 1816 -GPVSMAPTA--ITEAGTQPKQSVQKGEFVLATSSGAGVFKMGKRGFQVSAV----- 1866
Qy 2144 TQDTRHHPPQLSAPLAPLAPLSPFCASCPLVLDLRPPSDLYLPPP-----DHGAP 2193
Db 1867 --DGAQKCKNSEDAKSVHRESSTSSSVLSSSPSTLVKPPENGITIPGISDVPEP 1924
Qy 2194 ARGSPHSGGKRSPEPNKTS-----VLGGGEDGIE-----PVSPPPEGMTE 2233

Db 1925 AHKTASEAKSDTGQPKVGRFQVTTTANKVGRFSVSKTEKIDTDTKKGEFVASPPFMDL 1984
Qy 2234 PGHSRSAYVLLYRDGQTEPSPRMGSKSPGNTSOP--PAFFSKLTESNSAMVKSQKQETNK 2292
Db 1985 EQAVLPAVIPPKEKP-ELSEPHLN-----GPSDDPEAAFLSRDVEDGSGSPHSPHQLSSK 2039
Qy 2293 KLANTHNRNEPEYNISQPGTEIFNNPAITGTLMTYRSQAVQ-----BHASTNMGLEA 2344
Db 2040 SL-----PSQNLQSLSNSFNSYMSNEDIEDLKLRLRLDRKHLKEIQDLQS 2092
Qy 2345 IIRKALMGKQDQWEESEPLSANAFNPLNASASUPAAMPITAADGRSHDTLTPSGGGGKAK 2404
Db 2093 RQKHEIESLTGLKGVPP-----AVIIPPAAPLS---GRRRRPTKS-----KGS 2133
Qy 2405 VSGRPSRKAASP-APGLASGRDPPSV-----SSVHSEGD 2438
Db 2134 KSSRSSSLGNKSPQLSGNLGQSAASVLPHQOQLHPGN 2172

RESULT 94

Q20007 PRELIMINARY; PRT; 1274 AA.
AC Q20007; 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE F35A5.1 protein.
GN F35A5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodirinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
Smailson J., Smith A., Sonhammer E., Staden K., Sulston J.,
Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Leimbach D.;
RA "The sequence of C. elegans cosmid F35A5.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46675; AAB52641.1; -.
DR PIR; T16251; T16251.
DR WormPep; F35A5.1; CE04485.
SQ SEQUENCE 1274 AA; 138065 MW; 07F6B0D292C1799F CRC64;

Query Match 3.1%; Score 404; DB 5; Length 1274;
Best Local Similarity 20.1%; Pred. No. 3.4e-11;
Matches 353; Conservative 167; Mismatches 601; Indels 632; Gaps 87;
Qy 682 KKKKAPAAEEAFAFPVVEDEEAEASGVSNBEEVVEAEALHASGNEVPRGCSGPAT 741
Db 14 KKKKPP-----WESVDEEEVEVD-----EETPA-----PSKLEKPPSL 47


```
QY 2226 SPPEGMTEPGHRSNAVYLLYRDGROTEP-----SRMGSKSPGNTSOPPAFFSKLITESN 2279
Db 2573 HSPGAVAGVGTGAPVTWASSWATERAQOQRHSSATGRHPQRTVSP-----WRDLDN 2628
QY 2280 SAMVSKKQEQKINKLTHNRNEPEYNIQSGTEIFNMP--AITG 2321
Db 2629 SSPVR-----PRWDISQPSPIYHPAPGAITG 2655

RESULT 97
Q9VR13 ID Q9VR13 PRELIMINARY; PRT; 1118 AA.
AC Q9VR13;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG15635 protein.
GN CG15635.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dourbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fobel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon D., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
```



```
Db 283 -----AGFKGPGSGENGAPGQMGPRGLFCERGRGAP-----GPAGAR 322
Qy 974 PPIQVTKVHEPPREDAAATKPA--PPAPPPQNLQDESAPQOPGSS--PRGKSRSPAPP- 1029
Db 323 GNDGATGAAGPP-----GPTGPAAGPGPGAGCAKGEA--GPQPRGSEPGVGRGPGPG 377
Qy 1030 -----ADKEAFAAEAKLPGDPPCWTSLG--FPVPPREVIVKASHPADPSAFS 1076
Db 378 PAGAAGPAGNPGADQPGAKGANGAPG-----IAGAPGPGARGPGPGPGPGPKGN 432
Qy 1077 YAPPGHPLPLGHDTPARVLPBPPTISNPPPLISSAKHPSVLEROIGALSQMSVOLHVP 1136
Db 433 SGEPGAPSGK--DTGAKGEPGVGVQGGP-----GPAGEGRKG----- 470
Qy 1137 YSEHAKAPVGVMTGLPLPMDPKLAPFSGVKQEQLSRGOAGRPESLGVPVTAQEAIVLR 1196
Db 471 -----ARGEPGPT--GLP-----GPPGERGPGSGRPGADGVAGPK 505
Qy 1197 GTA--LGSPVGGSIKGIIP--STRVPSDSAITVRSITHGTADVLTKGTTITRIIGEDSPS 1253
Db 506 GPAGERGS--PGPAGPKGSPGEAGRPGEAGLPGAKGLT--GSPGSGPDGK----TGPPGPA 559
Qy 1254 RLD-----RGREDSL-----PKGHVYVEGKHVLSYEGM-----SVTQCKSED 1293
Db 560 QODGRPGPPPGARGQAGVMGFPKGAAEPGKAG-----ERGVPGPPGAVGAPGAKDG 614
Qy 1294 GRSSSGPPHETAAKRTYDMMGRVGRALSSAIEGLMGRAPP--ERHSPHILKXQHIIR 1352
Db 615 EAGAQQPP--GPAGPAGE-----RGEQGA--GSPGFQGLPGPAGPPGGEAKPGGE----- 660
Qy 1353 GSITGIGIPRSYVEAQEDYLREAKLLKREGTPPPPPSRDLTEAYKTQALGFLKLK--PA 1410
Db 661 -----QGVPGDLGAPGSPGARGERGPPGPGVQPGPA-----GPGANGAPG 704
Qy 1411 HGLVATVKEAGRSIHEIPRELRHLPBLAPRLKSGSIQGGP-----LKYDTGASTT 1466
Db 705 NDGAKG--DAG-----APGAP-----GS--QGAPGLQGMGERGAAGL 738
Qy 1467 GSKHD-----VRSILGSPGR-----TFP--PVHPLDVMADARALERACEYESLKRPG 1513
Db 739 PGPKGDRDAGKADGAGPGKGVGGLGTGPIGPPGAPGAPG-----KGSG 785
Qy 1514 TASSSGGTARGAPVIVPELAKPROSLTYEDHGAPFAGHLPRGSPVTMRPTPLRQGS 1573
Db 786 PSGPAGTCARGAPGDRGEGPPP--GPAGFA--GPPGAD-----GQPGAKGPGDAGAKGD 837
Qy 1574 LSSSKASQDRKLTSTPREIAKSPHSTVPEHHHPPI-----SPYEHLLRGVSGVDLYRSHIP 1629
Db 838 -----AGPPGAPGAPGPPGPIGNVGPAGKAGKAGSAGP----- 870
Qy 1630 LAFDPTSIPTGIPLDAAYLYLPHLANPTVPHLYPPYLRYGPDYTALENRQTIINDY 1689
Db 871 -----PGATGPPCAAGRVGP-----PGPS--GNAGPP-----GPPGPAKGKGGKPGRE- 912
Qy 1690 ITSQMHNTATAMAQRADML-----RGLSPRESSIALNYAAGPRGIIIDLSQVP 1738
Db 913 -----TGPAGRGVEGPPGPPGPGPAGEKSGPGADGAPAGCTPGPGIAGQGVV 961
Qy 1739 HLPVLVPPPT--PGTPATAMDLAYLPTAPQPPSSRHSSPSLSPGPGTHLTPTTTSSSR 1796
Db 962 GLPGQGRGERGPGPLPGPSGE-----PGKQPGSASGERGPPGPMGPPGLAGPGSGSRE- 1015
Qy 1797 ERDRERDRDREREKSIILTTTVEHAPIWRPGTEQSSGSGSGSGGG-----GSSSRP 1851
Db 1016 -----GAPGAGSGPGRDGSFGAKGDRGTGTPAGPP 1045
Qy 1852 ASHSHAHQHSPTSPRTQDALQORPSVLHNTGMKIIITAVEPSKPTVLRSSTSTSSVRPAA 1911
Db 1046 GAPGAPGAPGVGP-----AGKSGDRGTGTPAGP-----AGPVGPG 1082
Qy 1912 TFPPTHCPGLGTLGVYFTLMEPVLLPKEAPRVARPERPRADTGH----- 1957

Db 1083 ARGP-----AGQOPRGDKGETGEQDGRG1KGHRG 1112
Qy 1958 -AFLAKPPARGLEPASSPSKGSPPRLPLVPVSGHAT1ARTPAKNLA PHHASPDPAPPA 2016
Db 1113 FSGLGPPPCPPGSPGEGQPGSGASGP-----AGRGPPG 1145
Qy 2017 SASDPHREKTKOSKPSIOELRLSLGYHGSYSPEGVSPVSPSPSLTHDKGLPKHLEE 2076
Db 1146 SAGAPGKDLNGLPGPI-----GPPGPRGTGADGVGPPGPPG----- 1184
Qy 2077 LDKSHLEGELEPKPGPVKLGEEAHLPHLAPLPEBQSSSSPLLOTAPGVKGHQ----- 2130
Db 1185 -----PGPPGPPSAGDFSFPL-----QP-----POEKAHDGGRYR 1217
Qy 2131 -----RVVTLAQHISEVITODYTRHHPOQLSAPLAPLPSPPGASCPVL 2174
Db 1218 ADDANVVRDRDLEVDVTTLKSLSQIENIRSEGSKNPAR-----TC--R 1260
Qy 2175 DLRRPPSLYLPDPDHGAPARSPHSEGGKSPENKTSVLGG-----GEDGIEPV 2225
Db 1261 DLKMKCHSDW-----KSGEYWIDPNOGCLNDAIKVFCNMETGETCVYPT 1303
Qy 2226 SPPEGM-----TBPCHSRSAVYPLLYDGDQTEPSRMGSKSPGNTSOPPAFFSKLTSEN 2279
Db 1304 QPSVAQKNWYISKNPDKRHYVMFGESMTDGFQFEYG-----GQSGDPADVAIQLT--- 1353
Qy 2280 SAMVSKKQKQINKKLNTHNRNEPEYNISQPG 2310
Db 1354 --FLRLMSTEASQNTYTHCKNSVAYMDQQTG 1382

RESULT 100
QSNKP7
ID Q9NKP7 PRELIMINARY; PRT; 2936 AA.
AC Q9NKP7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE L712.2.
GN L712.2.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Worthey E.A., Slek E., Hixson G., Kiser P., Rickel E., Hassebrock M.,
RA Cawthra J., Sunkin S., Stuart K.D., Myler P.J.;
RT "Direct Submission."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RD EMBL: AC125735; AAMG9047.1;
SQ SEQUENCE 2936 AA; 305212 MW; BE689E280242FE6B CRC64;

Query Match 3.0%; Score 396.5; DB 5; Length 2936;
Best Local Similarity 19.4%; Pred. No. 2.3e-10;
Matches 462; Conservative 266; Mismatches 959; Indels 689; Gaps 101;

Qy 339 EIRKQRELOER-----MQSRVQORSG--L 361
Db 897 ELSQRLLRVCRDVSCMSAPAHAAVRARLASPTVSSAAARQPSLEFSTGASGGAGT 956
Qy 362 SMSAARSHEVSEIIDGLSEGENLEKOMRQLAVIPMLYDADQQRKFI--NMNGLMADPM 420
Db 957 SPSAAATHNW-----QVQHEEDQQLLFASIP-----RPAQAALKAVGSLRGAAADPS 1004
Qy 421 K-----VYKDRQVMNWSQEKETTFREKFMQHPKFNGLIASFLERKT---VA 464
Db 1005 RPLPPHFAVEARPLYPAQDPDASTTSGAAAPREPYE-----AALRELAERETHVVS 1058
Qy 465 EC-----VLYYYLTCKENYKSLVRSYRRRSKSSQOQQO-----QOQQOQQOQQO 509
Db 1059 ELDRLTMYDQLOEERRRQQRHRYAELLQSQQHKMLRHHRETVHLLQRECLQOQQO 1118
```


Job time : 398 secs

This Page Blank (uspto)